Geostatistical Mapping of Precipitation from Rain Gauge Data Using Atmospheric and Terrain Characteristics

PHAEDON C. KYRIAKIDIS,* JINWON KIM, AND NORMAN L. MILLER

Regional Climate Center, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California

ABSTRACT

A geostatistical framework for integrating lower-atmosphere state variables and terrain characteristics into the spatial interpolation of rainfall is presented. Lower-atmosphere state variables considered are specific humidity and wind, derived from an assimilated data product from the National Centers for Environmental Prediction and the National Center for Atmospheric Research (NCEP–NCAR reanalysis). These variables, along with terrain elevation and its gradient from a 1-km-resolution digital elevation model, are used for constructing additional rainfall predictors, such as the amount of moisture subject to orographic lifting; these latter predictors quantify the interaction of lower-atmosphere characteristics with local terrain. A "first-guess" field of precipitation estimates is constructed via a multiple regression model using collocated rain gauge observations and rainfall predictors. The final map of rainfall estimates is derived by adding to this initial field a field of spatially interpolated residuals, which accounts for local deviations from the regression-based first-guess field. Several forms of spatial interpolation (kriging), which differ in the degree of complexity of the first-guess field, are considered for mapping the seasonal average of daily precipitation for the period from 1 November 1981 to 31 January 1982 over a region in northern California at 1-km resolution. The different interpolation schemes are compared in terms of cross-validation statistics and the spatial characteristics of cross-validation errors. The results indicate that integration of low-atmosphere and terrain information in a geostatistical framework could lead to more accurate representations of the spatial distribution of rainfall than those found in traditional analyses based only on rain gauge data. The magnitude of this latter improvement, however, would depend on the density of the rain gauge stations, on the spatial variability of the precipitation field, and on the degree of correlation between rainfall and its predictors.

1. Introduction

Estimates of precipitation at regional scales are at the heart of hydrological, ecological, and environmental modeling in general, because they serve as input parameters (forcing fields) in spatially distributed models (Entekhabi et al. 1999). At these scales, the utility of precipitation predictions provided by general circulation models is limited because of their coarse spatial resolution. The use of limited area models (LAMs) is gradually emerging as a means for enhancing the accuracy of rainfall predictions at regional scales (Giorgi and Mearns 1991; Kim and Soong 1996; Miller and Kim 1996; Leung et al. 1996; Kim et al. 1998, 2000). Dynamic downscaling using LAMs yields multiple relevant variables, including precipitation, that are physically

and dynamically consistent. However, dynamic downscaling is computationally expensive and is not error-free, because of limited spatial resolution and model parameterizations. Statistical interpolation of rainfall based on rain gauge data still provides one of the basic analysis tools for constructing rainfall maps at regional scales (Tabios and Salas 1985; Dirks et al. 1998), even though physical and dynamic consistency of such interpolation predictions is not preserved in general.

In the context of mapping precipitation using rain gauge data, the variable most frequently used for enhancing interpolation, especially over mountainous regions, is terrain elevation (Chua and Bras 1982; Hevesi et al. 1992; Goovaerts 2000). Terrain-derived characteristics, such as slope and aspect, as well as other variables such as latitude, longitude, and distance from the coast, are less frequently accounted for in the mapping of rainfall (Spreen 1947; Burns 1953; Wolfson 1975; Wotling et al. 2000). Hybrid approaches also exist that account for slope orientation when selecting data to construct local regression models between rainfall and elevation (Daly et al. 1994). Although the above variables enhance the predictability of precipitation spatial dis-

*Current affiliation: Department of Geography, University of California, Santa Barbara, Santa Barbara, California.

Corresponding author address: Phaedon C. Kyriakidis, Dept. of Geography, University of California, Santa Barbara, Ellison Hall 5710, Santa Barbara, CA, 93106-4060.
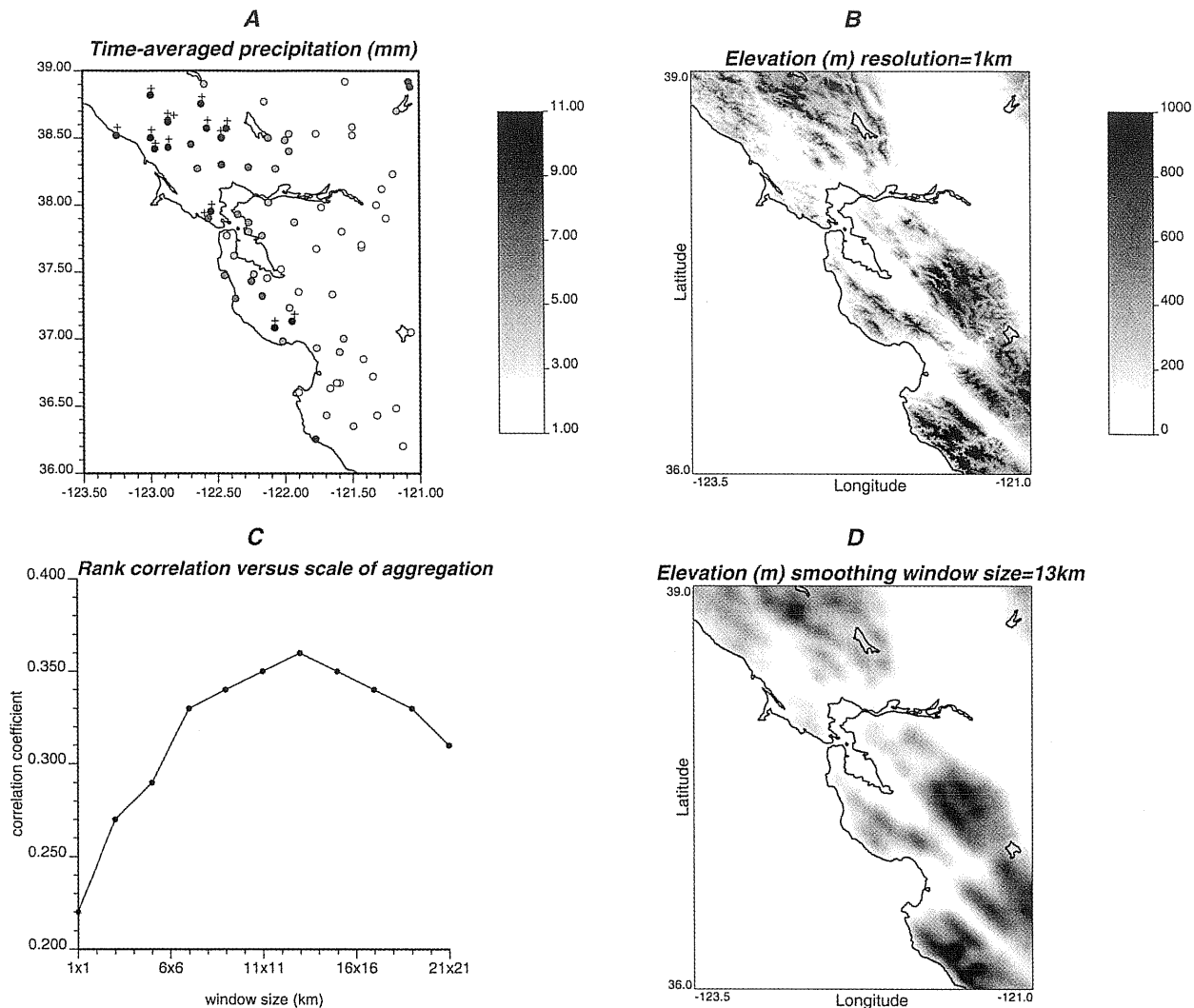E-mail: phaedon@geog.ucsb.edu

FIG. 1. Interaction of precipitation with elevation: (a) precipitation average (mm) during 1 Nov 1981–31 Jan 1982 at 77 rain gauges near the northern California coastal region (crosses indicate gauges used for jackknife), (b) digital elevation model of 1-km resolution (elevation values lower than 0 m are colored white, elevation values greater than 1000 m are colored black), (c) rank-order correlation coefficients between 77 collocated precipitation and window-averaged elevation values as a function of aggregation scale, and (d) smoothed elevation map derived via averaging the original elevation values of (b) within a 13 km × 13 km moving window.

tribution, they do not account directly for orographically induced rainfall (Smith 1979).

Simple regional models with limited atmospheric physics and dynamics [see Barros and Lettenmaier (1993) for a review] focus on lower-atmosphere state variables that steer storms and dictate the pattern and movement of air masses (Rhea 1978; Alpert 1986; Isakson 1996). Examples of such variables include wind speed and direction, as well as specific humidity integrated over several pressure levels (Pandey et al. 1999). Atmospheric variables typically are available at a very coarse resolution. They only provide a picture of the large-scale state of the atmosphere, which is expected to bear some relevance to observed precipitation at the local scale. The important link of such lower-atmosphere characteristics with precipitation lies in their in-

teraction with local terrain (Alpert and Shafir 1989b; Sinclair 1994; Andrieu et al. 1996).

To the authors' knowledge, apart from some aspects of the work of Herman et al. (1997), no comprehensive method has been previously reported for incorporating the joint effects of atmospheric and terrain variables to the spatial prediction of rainfall using geostatistical techniques. It should be stressed that such atmospheric variables are widely available at coarse resolutions, and their relevance to mapping precipitation at smaller scales is significant, as will be illustrated in what follows. The geostatistical framework presented in this paper accounts (in a quantitative way) for the joint effects of atmospheric and terrain variables into the mapping of rainfall at regional scales. This work is concerned with time-averaged precipitation; the additional factor

TABLE 1. Summary statistics (mm) of the 77 sample precipitation data.

Mean	Median	Std dev	Min	Max	Upper quartile	Lower quartile
5.83	5.22	3.03	1.49	14.35	8.01	3.40

TABLE 2. Matrix of Pearson correlation coefficient values between sample precipitation and its predictors ρ_{ZY_k}, $k = 1, \ldots, 7$ (top row) and between any two pairs of predictors, $\rho_{Y_kY_{k'}}$, k, $k' = 1, \ldots, 7$. The latter correlation values are computed over the entire study region (not only from the 77 rain gauges). The original predictors are specific humidity Y_1, elevation Y_2, and vertical wind Y_3. The interactions are humidity with elevation Y_4, humidity with vertical wind Y_5, elevation with vertical wind Y_6, and humidity with elevation and vertical wind Y_7.

	Z	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
Z	1.00	0.54	0.35	0.05	0.69	0.30	0.19	0.45
Y_1	—	1.00	−0.17	−0.02	0.68	0.68	−0.06	0.53
Y_2	—	—	1.00	0.03	0.48	−0.25	0.57	0.31
Y_3	—	—	—	1.00	0.03	0.62	0.76	0.61
Y_4	—	—	—	—	1.00	0.43	0.38	0.75
Y_5	—	—	—	—	—	1.00	0.37	0.77
Y_6	—	—	—	—	—	—	1.00	0.76
Y_7	—	—	—	—	—	—	—	1.00

of temporal variability is not addressed. Enhanced mapping of precipitation in space and time using atmospheric and terrain characteristics will be reported in the near future.

The selection of rainfall predictors based on physical considerations is demonstrated in section 2 for a study region in northern California. In section 3, alternative geostatistical approaches for mapping rainfall are presented. Two general cases are distinguished: spatial interpolation using (a) only rain gauge precipitation data and (b) rain gauge data in addition to low-atmosphere variables and their interaction with terrain. In section 4, a case study is undertaken: seasonal precipitation for the winter of 1981–82 is mapped for a region of northern California, using the various algorithms presented in section 3; the results are compared in terms of cross-validation statistics. Last, in section 5, some conclusions are drawn regarding the applicability of the proposed methodology to rainfall mapping and the potential avenues for future research.

2. Study area, precipitation, and its predictors

The study domain (Fig. 1a) is a 300 × 360 km² area of the northern California coastal region, which is characterized by complex terrain and extreme seasonal variation in precipitation. The characteristic length scale of the terrain ranges from approximately 50–100 km in the northern part of the domain (the Coast Range north of San Francisco Bay) down to 10–20 km in the south of the bay. Annual precipitation varies widely within the region from 200 mm yr⁻¹ in the Central Valley (east of the Coast Range) to over 1300 mm yr⁻¹ in the Santa Cruz Mountains (north of the Monterey Bay). Western slopes of the Coast Range receive 4–5 times more precipitation than the Central Valley during the cold season (November–March). Precipitation in the region is generally from stratiform clouds caused by orographic lifting of the westerly flow over the western slope of the Coast Range. On occasion, strong convection embedded within the stratiform clouds generates intense local precipitation.

The rainfall dataset used in this study consists of 77 rain gauge precipitation measurements representing the seasonal [November–December–January (NDJ)] average of daily rainfall for 1 November 1981–31 January 1982 at 77 stations over the study area (see Fig. 1a). The crosses attached to certain rain gauges indicate that these gauges are used subsequently in a jackknife procedure (see section 4). The statistics of the available precipitation data are shown in Table 1. Precipitation

values range from 1.49 to 14.35 mm with a mean of 5.83 mm and a standard deviation of 3.03 mm. The difference between the median (5.22 mm) and the mean (5.83 mm) values indicates a slightly positively skewed precipitation histogram. The original daily precipitation values constitute a subset of the Cooperative Observer and first-order precipitation stations, obtained from the National Oceanic and Atmospheric Administration; for details see Pandey et al. (1999).

The objective of this study is to map the season-average precipitation on a 300 × 360 grid of cell size 1 km², using all relevant information available for this region. The decision to map the particular NDJ average of precipitation, instead of an interannual mean, was dictated by the fact that such a map should be used as input to coupled hydrologic models calibrated for that particular time period. Mapping the interannual rainfall average might reveal different structures of dependence between precipitation and its predictors; such an alternative scenario was not investigated in this work.

a. Elevation as a precipitation predictor

The spatial distribution of precipitation is heavily influenced by temperature, especially by its vertical lapse rate, which dictates the local level (height) and rate of condensation. In the absence of detailed (small-scale) temperature information, we use elevation as its surrogate, at least as a first approximation; this would be true for the case of a spatially constant lapse rate. Because we are interested in the spatial patterns of temperature rather than in its absolute magnitude (see below), this first approximation is adequate for all practical purposes. One alternatively could establish a regression function between coarse-resolution (large scale) temperature and fine-resolution (small scale) elevation information. The fine-resolution terrain then could be transformed into a regression-based temperature map, which would exhibit small-scale variations due to corresponding local terrain variations.
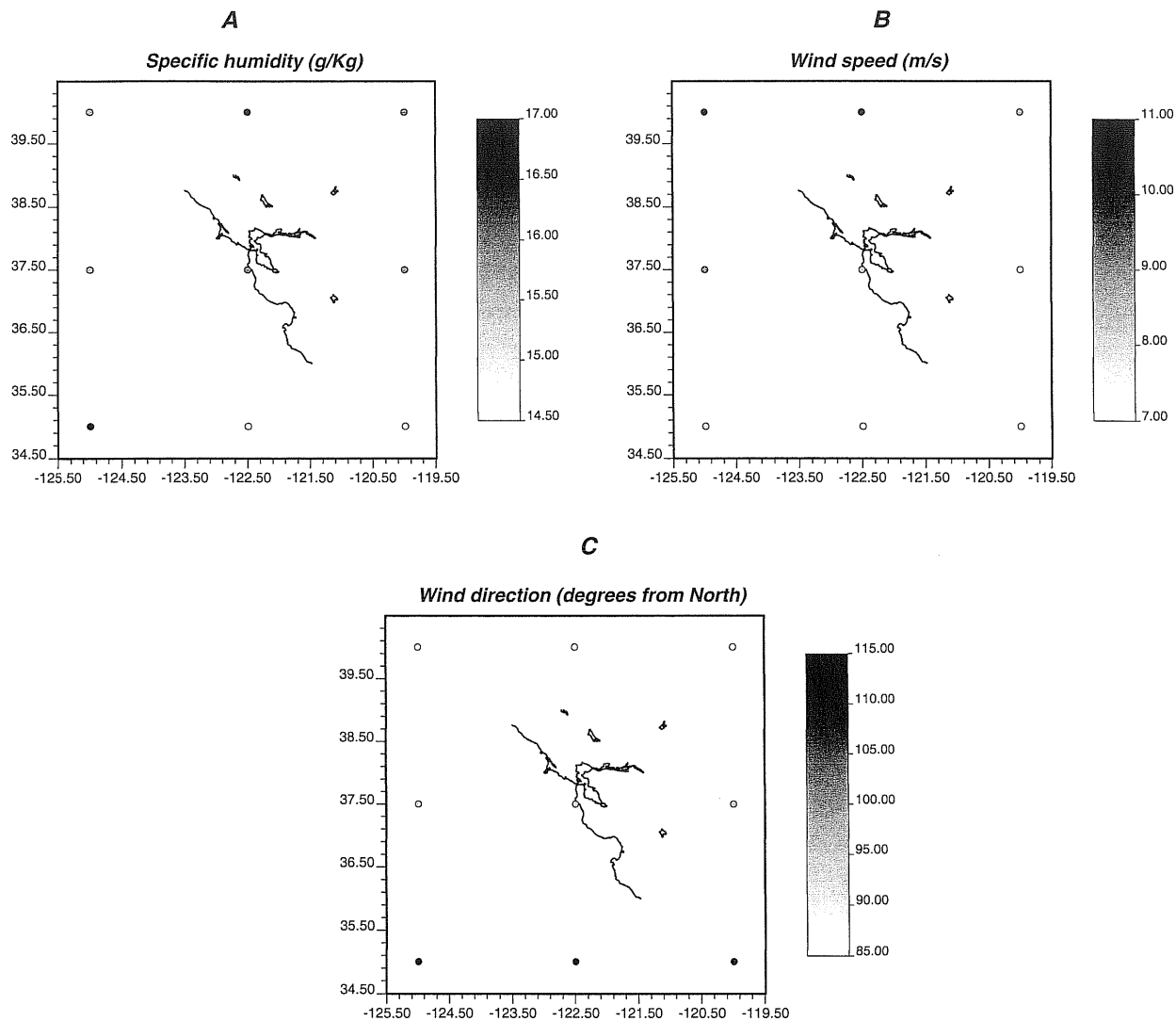
A

Specific humidity (g/Kg)

B

Wind speed (m/s)

C

Wind direction (degrees from North)

FIG. 2. Time-average (from 1 Nov 1981 to 31 Jan 1982) of daily lower-atmosphere state variables from NCEP–NCAR reanalysis nodes: (a) specific humidity (g Kg^{-1}) integrated from 850- to 1000-hPa levels, (b) wind speed (m s^{-1}) at the 700-hPa level, and (c) wind direction (in degrees clockwise from north) at the 700-hPa level.

A 1-km-resolution digital elevation model (DEM) is available for this area and is shown in Fig. 1b; elevation values range from -17 to 1668 m, with a median of 126 m and an interquartile range of 398 m. The DEM grid size is 300×360 km^2 and coincides with the grid at which estimation–interpolation of rainfall will be performed. All subsequently derived rainfall predictors are available at this spatial resolution. The first step, then, is to determine precipitation values at the 77 grid nodes that are closest to the 77 rain gauges using the nearest-neighbor method. The rank order (Spearman) correlation coefficient between collocated precipitation and elevation values is 0.22 (here the term collocated connotes that the pairs of rainfall and elevation data used to compute that correlation are located on the same grid node, possibly after nearest-neighbor interpolation of the rain

gauge value). This valve implies that, for the particular area and season, elevation is weakly correlated with precipitation, which is expected for relatively low mountains such as those found in the study region. We use the Spearman correlation coefficient instead of the ordinary Pearson correlation coefficient because the former is more robust to outliers. In addition, we interpret the relevance of elevation with precipitation in terms of its correlation with the relative rank of precipitation and less in terms of its correlation with precipitation magnitude. The nonlinear rank-order transformation involved in calculating the Spearman correlation coefficient allows capturing nonlinearities in the elevation–precipitation relationship.

To determine the scale of interaction between precipitation and elevation, the DEM-reported value at each

1-km cell was replaced by the elevation average over gradually enlarged square windows ranging from 3×3 km^2 to 21×21 km^2, with an increment of 2 km in each direction. Ten different averaging windows were applied to obtain 10 sets of averaged elevation values derived at the 77 rain gauges. Rank correlation coefficients were then calculated between the 10 sets of collocated precipitation and average elevation values corresponding to each window size. As indicated by the scatterplot of Fig. 1c, correlation increases gradually as the size of the averaging window increases, and a maximum of 0.36 is reached for a 13×13-km^2 window; for larger window sizes, correlation drops gradually. Window averaging is similar to low-pass filtering with a boxcar window, which smooths elevation spatial variability below a given window size. For the particular region shown in Fig. 1a, elevation exhibits the maximum relevance to precipitation, that is, can better inform its mapping, when small-scale (<13 km) elevation features are suppressed. Henceforth, the term elevation is used for this 13×13 km^2 window averaged elevation, whose map is shown in Fig. 1d.

b. Atmospheric variables as precipitation predictors

Lower-atmosphere state variables at the study region for the winter of 1981/82 include specific humidity, integrated from 850- to 1000-hPa levels, and the horizontal wind components at the 700-hPa level. More specific, the time averages of specific humidity and horizontal wind components over the period of interest were retained. Lower-atmosphere state variables are available at a coarse resolution of $2.5° \times 2.5°$ from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis dataset (Kalnay et al. 1996). The reanalysis products provide snapshots of the global atmospheric and surface fields at a uniform resolution every 6 h. NCEP–NCAR reanalysis variables are generated via data assimilation, whereby observational data around the globe are used to drive a global atmospheric model. Details of the data quality control and assimilation procedure are presented in Kalnay et al. (1996). Together with the reanalysis from the European Centre for Medium-range Weather Forecasts, NCEP–NCAR reanalysis variables are regarded as the most reliable (and widely available) representation of the instantaneous state of the global atmosphere.

We retained reported values at nine reanalysis nodes within and nearby the study region, which are shown in Fig. 2. Season-averaged specific humidity ranges from 14.81 to 16.68 g kg^{-1} and is relatively smaller in the southeastern part of the study domain (Fig. 2a). Wind speed ranges from 7.02 to 10.90 m s^{-1} and is relatively higher at the northwestern part of the region (Fig. 2b). Last, wind direction ranges from 86.58° to 110.61° from the north (Fig. 2c), indicating a dominant wind direction from the west. Lower-atmosphere state

variables, although known at a very small number of locations, provide a picture of the large-scale state of the lower atmosphere, which is expected to be related to observed precipitation at the local (1 km) scale.

We use interpolated horizontal wind components (not shown) to calculate the vertical component of wind w_s due to orographic lifting at the local scale. This terrain-induced vertical motion is defined as the inner product of directional elevation gradients with the corresponding horizontal wind components, (e.g., Alpert 1986):

$$w_s = -\mathbf{v} \cdot \nabla h = -\left[u \left(\frac{dh}{dx} \right) + v \left(\frac{dh}{dy} \right) \right], \quad (1)$$

where $\mathbf{v} = (u, v)$ denotes the wind vector with local horizontal components u and v, and dh/dx, dh/dy denote the local gradients of elevation along the same directions.

Although the horizontal wind vector $\mathbf{v}$ is available at a very coarse resolution from the nine NCEP nodes and interpolation at a 1-km resolution provides very smooth maps of the horizontal wind components over the study region, the vertical wind component due to orographic lifting exhibits small-scale variations due to local terrain. In this paper, such terrain-induced vertical motion, which can also be regarded as the exposure of local terrain to local wind, is accounted for in the regional-scale mapping of rainfall.

Rainfall predictors are subsequently derived from the NCEP–NCAR reanalysis lower-atmosphere state variables and their interaction with local terrain in the following three steps.

1) Interpolation (via the inverse-distance-squared method) of specific humidity and horizontal wind components from the nine NCEP nodes to a 300×360 grid with 1-km^2 cell size is done. Inverse-distance interpolation is retained in lieu of other methods of objective analysis because of the very small number of NCEP nodes, which does not allow for reliable inference of structural characteristics, for example, a spatial correlation model for specific humidity. The vertical wind component at any grid cell is calculated using Eq. (1). Let $x_1(\mathbf{u})$, $x_3(\mathbf{u})$ denote the interpolated values of specific humidity and derived vertical wind component at any grid cell with (2D) coordinate vector $\mathbf{u} = (u_1, u_2)$, expressed, for example, in degrees of longitude and degrees of latitude, respectively. Let $x_2(\mathbf{u})$ denote the elevation at $\mathbf{u}$ obtained from the low-pass filtering procedure described in section 2a.

2) The resulting values are transformed to follow a uniform distribution. For the case of specific humidity, for example, the transformed value $y_1(\mathbf{u})$ at any location $\mathbf{u}$ is computed as

$$y_1(\mathbf{u}) = F_{X_1}[x_1(\mathbf{u})], \quad \in [0, 1], \quad (2)$$

where $F_{X_1}(\ )$ denotes the cumulative histogram of the interpolated specific humidity values. For the case of vertical wind component X_3, a zero value corre-
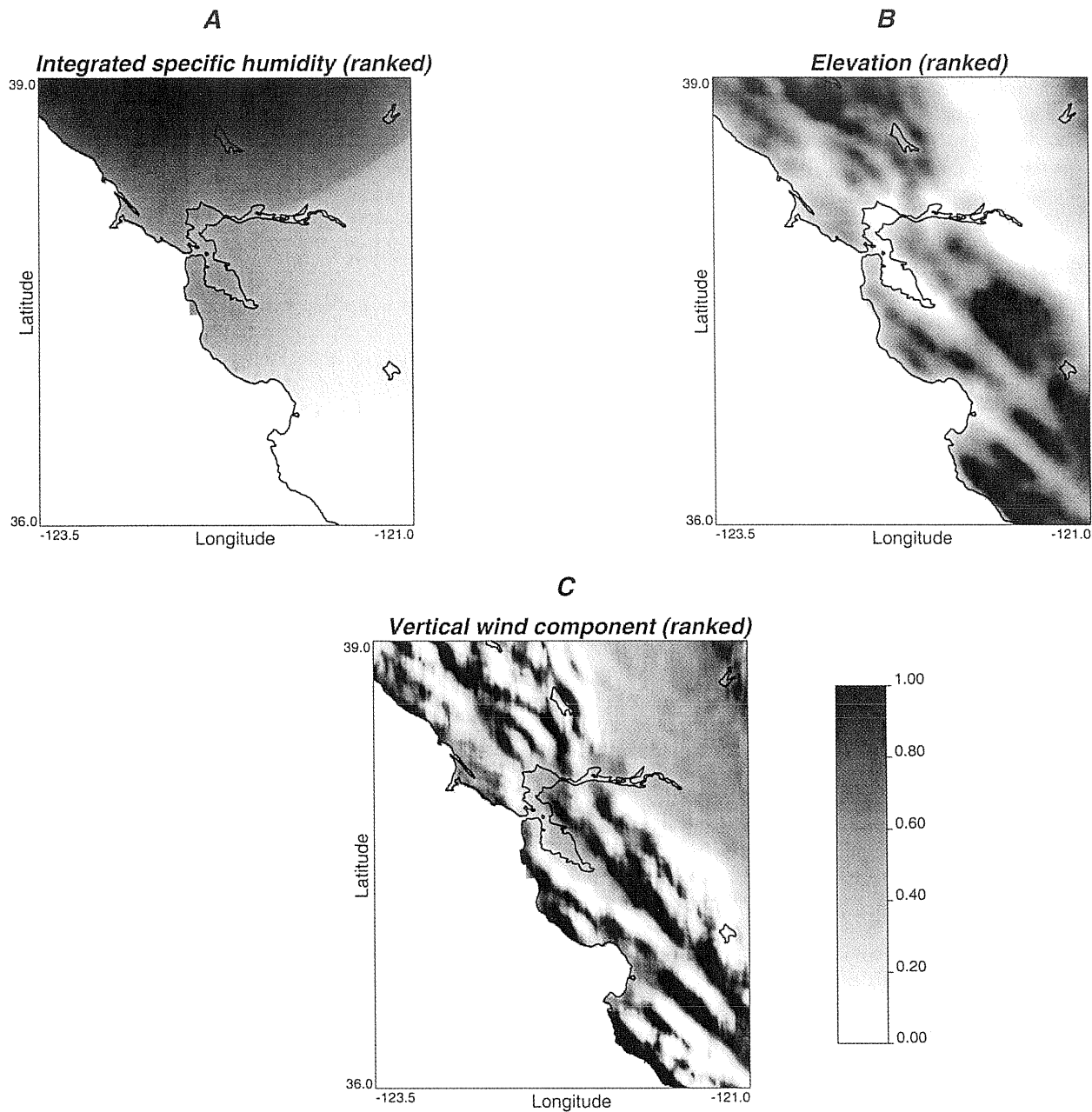
FIG. 3. Maps of ranked predictors for precipitation: (a) specific humidity integrated from 850 to 1000 hPa, (b) average elevation within a 13-km square window, and (c) vertical wind component.

sponds to the median of the distribution; that is, $F_{X_3}(0.0) = 0.5$, which implies that a rank-ordered value $y_3(\mathbf{u})$ of greater than 0.5 should be interpreted as uplift. In the more general case, one would first compute the probability corresponding to a zero value of the vertical wind component $p_3^{(0)} = F_{X_3}(0.0)$ and then interpret as uplift all rank-ordered values greater than $p_3^{(0)}$. Rank-ordered values lower than $p_3^{(0)}$ conversely would correspond to downslope vertical wind movement.

3) The interaction terms among the three predictors $y_1(\mathbf{u})$, $y_2(\mathbf{u})$, and $y_3(\mathbf{u})$ are calculated. For example, the interaction between humidity and the vertical wind component at a grid cell $\mathbf{u}$ is calculated as $y_5(\mathbf{u}) = y_1(\mathbf{u})y_3(\mathbf{u})$. This product term is the rank-ordered amount of moisture uplift due to wind encountering the local terrain slope. The interaction term $y_7(\mathbf{u}) = y_1(\mathbf{u})y_2(\mathbf{u})y_3(\mathbf{u})$ is the (rank ordered) moisture uplift due to wind encountering the local terrain slope, modulated by the decrease of moisture availability

TABLE 3. Model parameters adopted (via cross validation) for the variograms of the sample precipitation data, and the residuals from the different trend functions. The variogram model specification is $\gamma^{(i)}(|h|) = C^{(i)}(0) - c_1^{(i)} \exp[|h|/a_1^{(i)}] - c_2^{(i)} \exp[|h|/a_2^{(i)}]$, where $|h|$ denotes the modulus of a distance vector h, and $C_i(0)$ denotes the variance of the ith dataset. Parameters $c_1^{(i)}$ and $a_1^{(i)}$ denote the respective sill (mm²; contribution to the total variance of the ith dataset) and range (km) of the small-scale variogram structure; $c_2^{(i)}$ and $a_2^{(i)}$ denote the respective sill and range of the large-scale variogram structure.

Dataset	c_1	c_2	a_1	a_2
Original z values	2.00	7.16	40	160
Residuals REGR-2	0.57	7.50	40	150
Residuals REGR-4	1.00	3.80	30	120
Residuals REGR-347	1.00	3.35	25	90

with topographic height. The remaining interaction terms were calculated as $y_4(u) = y_1(u)y_2(u)$ and $y_6(u) = y_2(u)y_3(u)$. All the resulting interaction fields are also transformed to follow a uniform distribution as in Eq. (2).

The transformation procedure of Eq. (2) is applied for eliminating the effect of different histograms of predictor fields to precipitation mapping, given that all transformed fields lie in the interval [0, 1]. One should interpret the relevance of a predictor value at any grid cell in terms of its relation with the relative rank of precipitation rather than in terms of its correlation with precipitation magnitude. The magnitude of precipitation is dictated by the rain gauge observations themselves.

Maps of the final rank-transformed values of specific humidity $y_1(u)$, elevation $y_2(u)$, and vertical wind component $y_3(u)$ at any 1-km² grid cell u are shown in Fig. 3. Note the extremely smooth spatial variation of humidity (Fig. 3a), as opposed to that of elevation (Fig. 3b) and vertical wind component (Fig. 3c). The patterns of spatial variability of the latter are mainly due to the elevation gradient fields, because the interpolated horizontal wind components (not shown) exhibit very smooth spatial variation attributable to the limited spatial resolution of NCEP–NCAR reanalysis. The vertical wind component field (Fig. 3c) exhibits the most complex spatial variability. Relatively large positive values of vertical wind, dark pixels corresponding to stronger uplift, are found on the windward side of the terrain (recall that the prevailing wind direction is from the west).

Maps of the final rank-transformed values of the in-

teractions between humidity and elevation $y_4(u)$, humidity and vertical wind $y_5(u)$, elevation and vertical wind $y_6(u)$, and humidity with elevation and vertical wind $y_7(u)$ are shown in Fig. 4. One can appreciate the smooth spatial patterns inherited from humidity and to a lesser extent from elevation (Fig. 4a) and the complex spatial patterns inherited from the vertical wind component (Figs. 4b–d).

Last, we investigate whether the predictor values at the rain gauge stations are a representative sample of their respective populations. This is done by constructing quantile–quantile plots between the distributions of all predictors (over the entire domain) and their respective sample distributions (Fig. 5). A representative sample for, say, elevation would lead to a corresponding quantile–quantile plot aligned with the 45° line. Rain gauges tend to be located in relatively low elevations, although very low and very high elevations are adequately covered (Fig. 5a). Specific humidity and vertical wind, on the other hand, are adequately sampled from the spatial configuration of the available rain gauges (Figs. 5b–c). Similar graphs were constructed for the remaining predictors (not shown), and indicated representative samples for these predictors.

c. Correlation between precipitation and its predictors

In this paper, the relevance of the three predictors and their interactions to observed precipitation is quantified in terms of their correlation with the rain gauge data (Table 2, top row). Maximum correlation between a predictor and precipitation is reached for the interaction of humidity with elevation ($\rho_{ZY_4} = 0.69$), followed by humidity alone ($\rho_{ZY_1} = 0.54$) and by the interaction of humidity with elevation and vertical wind component ($\rho_{ZY_7} = 0.45$). Elevation alone is not strongly correlated with precipitation ($\rho_{ZY_2} = 0.35$), and the same holds true for the interaction of humidity with vertical wind ($\rho_{ZY_5} = 0.30$). The vertical wind component and its interaction with elevation are almost uncorrelated with precipitation ($\rho_{ZY_3} = 0.05$ and $\rho_{ZY_6} = 0.19$, respectively). These low correlation-coefficient values might be the result of the small time averaging period (NDJ), which cannot reveal a definite correlation between vertical wind component and observed precipitation. Such a relationship could be more evident if observed precipi-

TABLE 4. Statistics of cross-validation errors for different mapping algorithms (subscripts denote the predictors used in the respective regression equations). Rmse denotes the root-mean-square error (mm), $\rho(Z_{cr}^*, Z)$ denotes the correlation between the cross-validation estimates $z_{cr}^*(u_\alpha)$ and the sample precipitation data $z(u_\alpha)$, and $\rho(E_{cr}^*, Z)$ denotes the correlation between the cross-validation errors $e_{cr}^*(u_\alpha)$ and the true values $z(u_\alpha)$. The numbers in parentheses show the relative changes in the corresponding cross-validation statistics from OK (values in the first column).

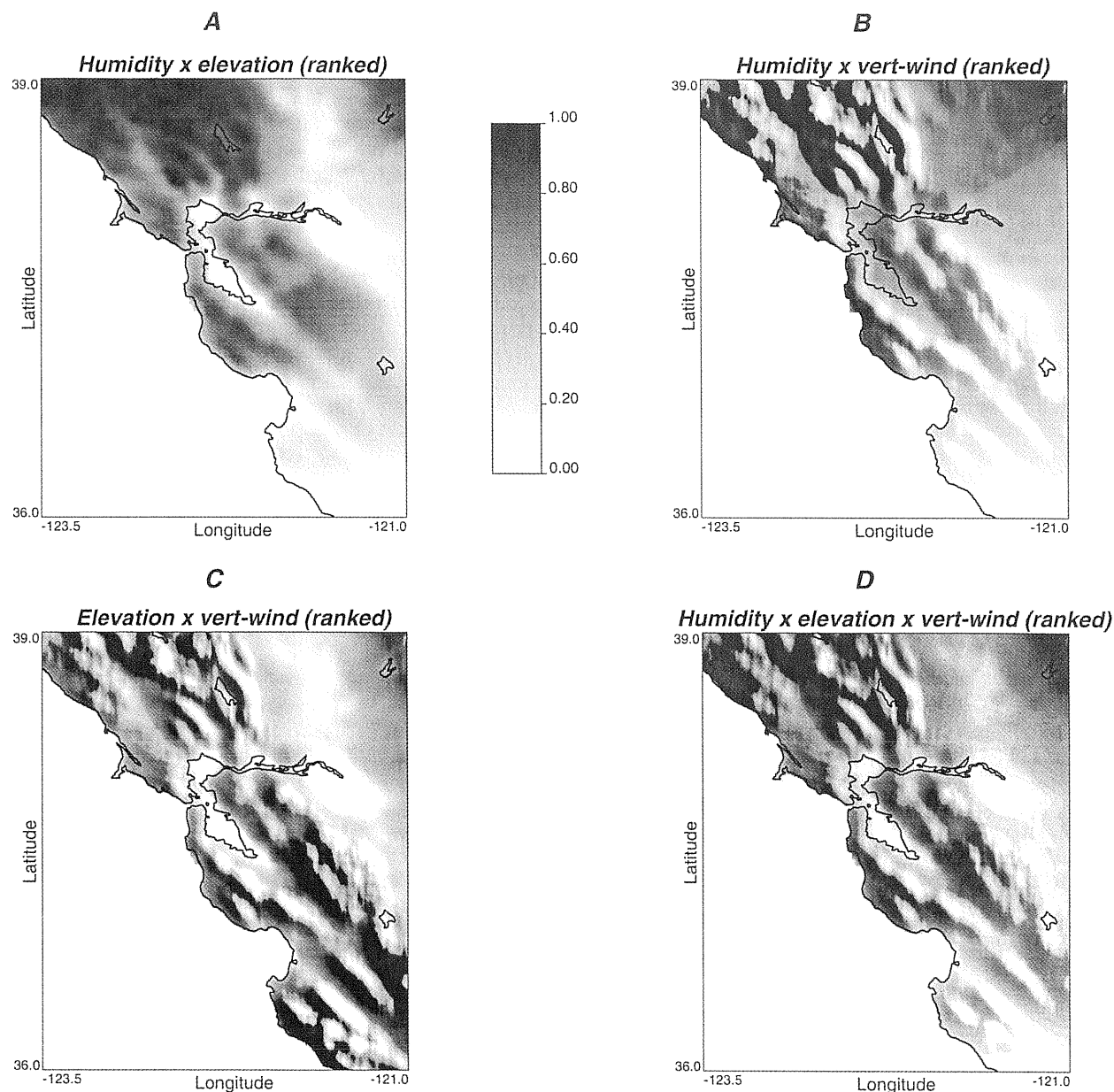| Statistics	OK	SKLM$_2$	KED$_2$	SKLM$_4$	KED$_4$	SKLM$_{347}$	KED$_{347}$
Rmse	1.67	1.59	1.58	1.66	1.64	1.56 (−7%)	1.52 (−9%)
$\rho_{Z_{cr}^* Z}$	0.83	0.86	0.85	0.84	0.84	0.86 (+4%)	0.87 (+5%)
$\rho_{E_{cr}^* Z}$	−0.54	−0.47	−0.44	−0.44	−0.48	−0.37 (−31%)	−0.46 (−15%)

FIG. 4. Maps of ranked interactions between precipitation predictors: (a) humidity with elevation; (b) humidity with vertical wind component, (c) elevation with vertical wind, and (d) humidity with elevation and vertical wind.

tation within different storm types (e.g., frontal vs synoptic) was compared with the vertical wind component occurring during the same storm type.

From Table 2, one can appreciate the importance of humidity in mapping precipitation, given that its (Pearson) correlation with the rain gauge data is $\rho_{ZY_1} = 0.54$. However, the spatial variation of humidity is very smooth and cannot distinguish local precipitation features. It is here where the role of the interaction between lower-atmosphere state variables and terrain characteristics becomes important. Indeed, the interaction of humidity with elevation is strongly correlated with pre-

cipitation ($\rho_{ZY_4} = 0.69$), and the interaction of humidity with elevation and vertical wind is moderately correlated with precipitation ($\rho_{ZY_7} = 0.45$). Both of these predictor fields exhibit nonsmooth patterns of spatial variability and potentially can resolve local precipitation features.

Next, we investigate whether a significant correlation exists between the predictor variables themselves, a situation termed mutlicollinearity in a regression context (Draper and Smith 1998). The original predictors (i.e., specific humidity Y_1, elevation Y_2, and the vertical wind component Y_3) constitute nearly independent variables,

A

Q-Qplot: Station vs pixel elevation

B

Q-Qplot: Station vs pixel specific humidity

C

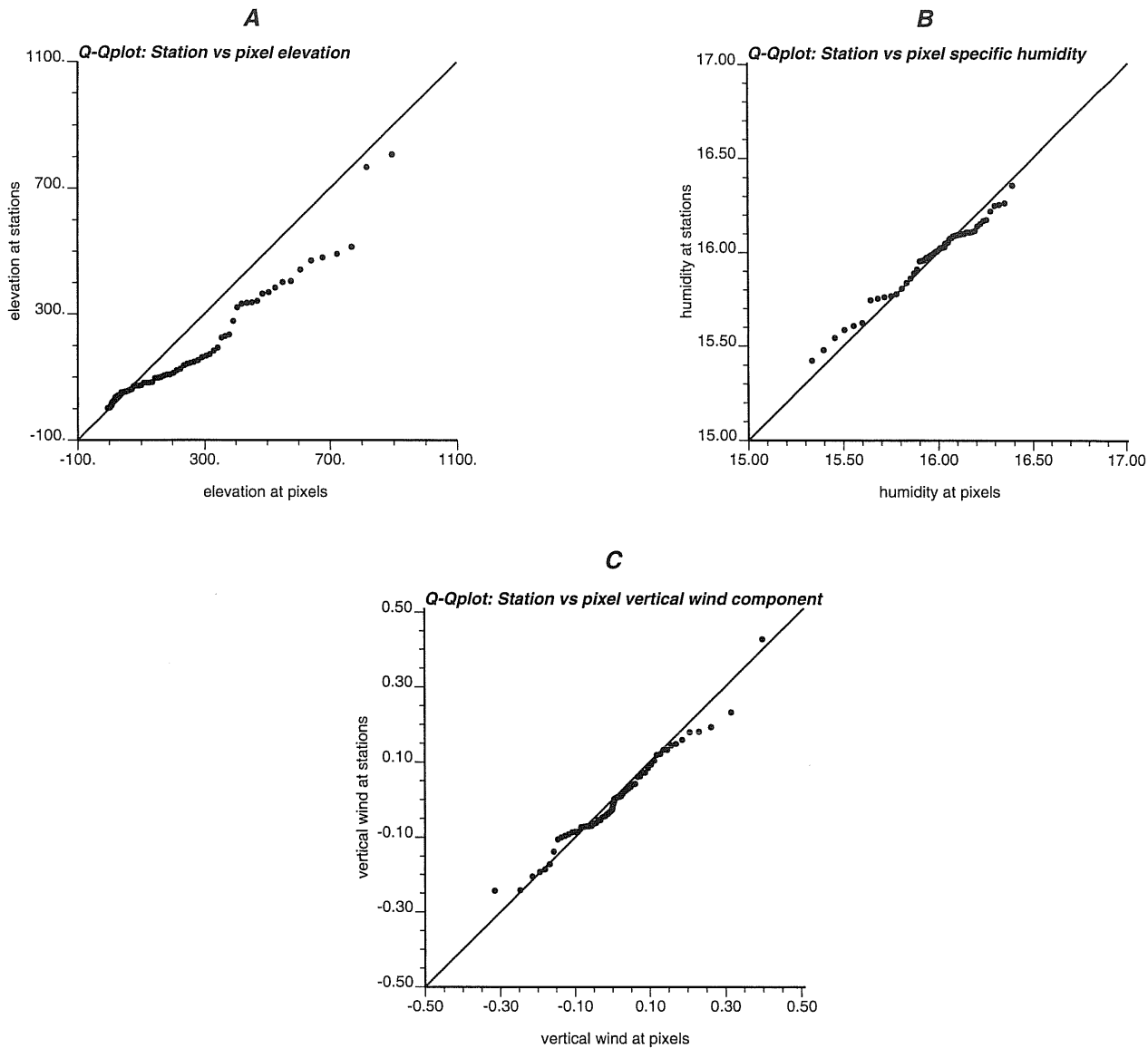**Q-Qplot: Station vs pixel vertical wind component**

FIG. 5. Plots to check whether predictor values at rain gauges constitute representative samples from their respective populations. Quantile–quantile plots between distributions of (a) gauge vs DEM reported elevations, (b) interpolated specific humidity at 77 pixels closest to the 77 rain gauges vs specific humidity at all pixels, and (c) vertical wind component at 77 pixels closest to the 77 rain gauges vs specific humidity at all pixels.

because the correlation between them is very low: $\rho_{Y_1 Y_2} = -0.17$, $\rho_{Y_1 Y_2} = -0.02$, and $\rho_{Y_2 Y_3} = 0.03$ (see Table 2). The interaction variables (i.e., that of specific humidity with elevation Y_4, specific humidity with vertical wind Y_5, elevation with vertical wind Y_6, and specific humidity with elevation and vertical wind Y_7) are not independent, as indicated by the nonnegligible correlation coefficients between them (Table 2). In a multiple regression model, interpretation of the resulting regression coefficients is straightforward only in the case of independent predictors. When predictors are correlated, as are the derived precipitation predictors (interactions) in this dataset, the resulting regression co-

efficients should be interpreted in the light of multicollinearity.

3. Mapping precipitation via geostatistics

Consider the task of predicting the unknown precipitation value $z(\mathbf{u})$ at any location $\mathbf{u}$ within the study area, that is, the task of constructing a map of precipitation estimates. Atmospheric and terrain variables, as well as their multiplicative effects (their interactions), constitute valuable information for improving precipitation predictability, in addition to the n available rain gauge measurements $\{z(\mathbf{u}_\alpha), \ \alpha = 1, \ldots, n\}$ (here $\mathbf{u}_\alpha$

denotes the coordinate vector of the αth rain gauge). The objective is (a) to assess the relevance of each predictor to observed precipitation and (b) to account for the most relevant predictors in the spatial interpolation of rainfall.

For simplicity, we assume that $\mathbf{u}$ represents the coordinate vector of the central point of any grid cell and that rain gauge data or relevant atmospheric variables (derived or interpolated) are defined on the same quasi-point support. In our case study (see section 4), an unknown precipitation value $z(\mathbf{u})$ or a known elevation value $y_2(\mathbf{u})$ is regarded as representative of a 1-km^2 grid cell centered at $\mathbf{u}$.

In geostatistics, a precipitation value $z(\mathbf{u})$ at a location $\mathbf{u}$ within the study area (whether a rain gauge or not) is regarded as an outcome of a random variable (RV) $Z(\mathbf{u})$ at that location. The set of all possible spatially dependent RVs in the study area constitutes a random function or random field (RF) denoted as $\{Z(\mathbf{u}), \mathbf{u} \in D\}$ (Journel and Huijbregts 1978; Goovaerts 1997; Chilès and Delfiner 1999). The RF $\{Z(\mathbf{u}), \mathbf{u} \in D\}$ is typically decomposed as

$$Z(\mathbf{u}) = m(\mathbf{u}) + R(\mathbf{u}), \qquad \mathbf{u} \in D, \qquad (3)$$

where $m(\mathbf{u})$ is a deterministic trend component modeling an "average" spatial variation of $Z(\mathbf{u})$ and $R(\mathbf{u})$ is a stochastic residual RF with zero mean $E[R(\mathbf{u})] = 0, \forall \mathbf{u}$ and covariance $C_R(\mathbf{u}, \mathbf{u}') = E[R(\mathbf{u})R(\mathbf{u}')]$. The expected value of $Z(\mathbf{u})$ at any location is therefore the trend value $m(\mathbf{u})$ at that location; that is, $E[Z(\mathbf{u})] = m(\mathbf{u}), \mathbf{u} \in D$.

The decomposition of Eq. (3) is entirely subjective, because no data exist on either $m(\mathbf{u})$ or $R(\mathbf{u})$ (Thiébaux 1997). In all rigor, Eq. (3) should be rewritten as

$$Z(\mathbf{u}) = m_{\mathrm{alg}}(\mathbf{u}) + R_{\mathrm{alg}}(\mathbf{u}), \qquad \mathbf{u} \in D, \qquad (4)$$

to emphasize the dependence of the trend and residual components to the particular algorithm (alg) used to define them.

On one hand, the trend component $m_{\mathrm{alg}}(\mathbf{u})$ could be identified to predictions of a physically based deterministic model, and the resulting residual $R_{\mathrm{alg}}(\mathbf{u})$ could be viewed as stochastic spatial variability (Rutherford 1972). On the other hand, a constant trend component $m_{\mathrm{alg}}(\mathbf{u}) = m$ attributes all spatial variability to the residual component $R_{\mathrm{alg}}(\mathbf{u})$, which implies that no prior knowledge exists regarding the average spatial variability of the phenomenon under study. In the optimal case, the trend component should be associated with some physically meaningful component of spatial variability. The residual component should model stochastic variations (not necessarily purely random, i.e., white noise) around that trend. In what follows, the subscript alg is dropped for simplicity.

If the spatial variation of precipitation is considered to be stationary, the expected value of any RV $Z(\mathbf{u})$ is equal to a constant m, independent of the coordinate vector $\mathbf{u}$:

$$m = m(\mathbf{u}) = E[Z(\mathbf{u})] \simeq \frac{1}{n} \sum_{\alpha=1}^{n} z(\mathbf{u}_\alpha), \qquad \forall \mathbf{u} \in D,$$

and the residual covariance function $C_R(\mathbf{u}, \mathbf{u}')$ between any two residual values $r(\mathbf{u})$ and $r(\mathbf{u}') = r(\mathbf{u} + \mathbf{h})$ depends only on the magnitude (and possibly the orientation) of the vector $\mathbf{h}$ separating them:

$$C_R(\mathbf{u}, \mathbf{u}') = C_R(\mathbf{h}) = E[R(\mathbf{u})R(\mathbf{u} + \mathbf{h})]$$

$$\simeq \frac{1}{n(\mathbf{h})} \sum_{\alpha=1}^{n(\mathbf{h})} r(\mathbf{u}_\alpha)r(\mathbf{u}_\alpha + \mathbf{h}),$$

where $n(\mathbf{h})$ denotes the number of residual pairs $\{r(\mathbf{u}_\alpha), r(\mathbf{u}_\alpha + \mathbf{h})\}$ separated by vector $\mathbf{h}$. The covariance $C_R(\mathbf{h})$ of the residual r values for the stationary case is equal to that of the original precipitation z values; that is, $C_R(\mathbf{h}) \simeq C(\mathbf{h})$.

Before presenting rainfall mapping schemes that account for several auxiliary variables, such as humidity, elevation, and their interactions, we first present the case of mapping rainfall using only rain gauge measurements.

a. Spatial interpolation using only rain gauge data

Consider the task of estimating the unknown precipitation value $z(\mathbf{u})$ at a location $\mathbf{u}$ from $n(\mathbf{u}) < n$ surrounding values $\{z(\mathbf{u}_\alpha), \alpha = 1, \ldots, n(\mathbf{u})\}$ within a neighborhood $W(\mathbf{u})$ centered at $\mathbf{u}$. In the stationary case, $n(\mathbf{u}) + 1$ pieces of information are available: the known regional mean m, and the $n(\mathbf{u})$ residual values $\{r(\mathbf{u}_\alpha) = [z(\mathbf{u}_\alpha) - m], \alpha = 1, \ldots, n(\mathbf{u})\}$.

The simple kriging (SK) estimate $z^*_{\mathrm{SK}}(\mathbf{u})$ for the unknown value $z(\mathbf{u})$ is expressed as

$$z^*_{\mathrm{SK}}(\mathbf{u}) = m(\mathbf{u}) + r^*_{\mathrm{SK}}(\mathbf{u}) = m + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})r(\mathbf{u}_\alpha), \quad (5)$$

where $r^*_{\mathrm{SK}}(\mathbf{u})$ denotes the SK estimate (with known zero mean) of the unknown residual $r(\mathbf{u})$ and $\lambda_\alpha(\mathbf{u})$ denotes the weight assigned to the αth residual value $r(\mathbf{u}_\alpha)$ when estimation is performed at location $\mathbf{u}$.

The $n(\mathbf{u})$ weights $\{\lambda_\alpha(\mathbf{u}), \alpha = 1, \ldots, n(\mathbf{u})\}$ are obtained per solution of the system of normal equations or SK system:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_R(\mathbf{u}_\alpha - \mathbf{u}),$$

$$\alpha = 1, \ldots, n(\mathbf{u}), \quad (6)$$

where $C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta)$ is the covariance between any two data locations $\mathbf{u}_\alpha$ and $\mathbf{u}_\beta$, and $C_R(\mathbf{u}_\alpha - \mathbf{u})$ is the covariance between any datum location $\mathbf{u}_\alpha$ and the location $\mathbf{u}$ where estimation is performed. Note that the weights do not depend on the residual values, only on their configuration through the covariance model $C_R(\mathbf{h})$. In meteorology, simple kriging is termed objective analysis (Daley 1991), and in the statistical literature it is termed best linear unbiased predictor (Cressie 1993).

In reality, the regional mean m of precipitation is unknown; it could be estimated by a weighted average of the sample rainfall data using various weighting schemes. In practice, however, the average precipitation spatial variability is nonstationary; that is, it varies systematically from one location to another. Local variations of the trend component m can be accounted for by considering a mean $m(\mathbf{u})$, which is constant within each neighborhood $W(\mathbf{u})$ but varies from one neighborhood to another:

$$m(\mathbf{u}') = c, \qquad \forall \mathbf{u}' \in W(\mathbf{u}),$$
$$\neq c' = m(\mathbf{u}'''), \qquad \forall \mathbf{u}''' \in W(\mathbf{u}'').$$

Estimation of the unknown value $z(\mathbf{u})$ now proceeds by first estimating the local mean $m(\mathbf{u})$ within the neighborhood $W(\mathbf{u})$, using data only from that neighborhood,

and then performing SK using the resulting residuals. The algorithm is known as ordinary kriging (OK) with moving neighborhoods (or moving windows).

Within any neighborhood $W(\mathbf{u})$, the OK estimate $z^*_{\text{OK}}(\mathbf{u})$ is written as

$$z^*_{\text{OK}}(\mathbf{u}) = m^*_{\text{OK}}(\mathbf{u}) + r'_{\text{SK}}(\mathbf{u})$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \nu_\alpha(\mathbf{u}) z(\mathbf{u}_\alpha) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) r'(\mathbf{u}_\alpha), \quad (7)$$

where $r'(\mathbf{u}_\alpha)$ denotes a new residual value defined as: $r'(\mathbf{u}_\alpha) = z(\mathbf{u}_\alpha) - m^*_{\text{OK}}(\mathbf{u}_\alpha)$.

The $n(\mathbf{u})$ weights $\{\nu_\alpha(\mathbf{u}), \alpha = 1, \ldots, n(\mathbf{u})\}$, which are different from the SK weights $\{\lambda_\alpha(\mathbf{u}), \alpha = 1, \ldots, n(\mathbf{u})\}$, hence the different notation, are obtained per solution of the OK system:

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \nu_\beta(\mathbf{u}) C_{R'}(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_{\text{OK}}(\mathbf{u}) = 0, \qquad \alpha = 1, \ldots, n(\mathbf{u}) \\ \\ \sum_{\beta=1}^{n(\mathbf{u})} \nu_\beta(\mathbf{u}) = 1, \end{cases} \quad (8)$$

where $C_{R'}(\mathbf{u}_\alpha - \mathbf{u}_\beta)$ denotes the covariance between any two residual values $r'(\mathbf{u}_\alpha)$ and $r'(\mathbf{u}_\beta)$, and $\mu_{\text{OK}}(\mathbf{u})$ is a Lagrange parameter due to the constraint on the weights imposed for estimating the local mean $m^*_{\text{OK}}(\mathbf{u})$. In this local stationary case, the typical practice consists of assimilating the covariance $C_{R'}(\mathbf{h})$ of the residuals to the covariance $C_R(\mathbf{h})$ (Goovaerts 1997). This is the reason for using the same notation for the SK weights $\lambda_\alpha(\mathbf{u})$ in both Eqs. (5) and (7).

Kriging (either SK or OK) is an exact interpolator; that is, kriging estimates reproduce observed sample data values at their locations: $z^*_{\text{OK}}(\mathbf{u}_\alpha) = z(\mathbf{u}_\alpha)$ for any sampling location $\mathbf{u}_\alpha$. This data-exactitude property of kriging is not shared by the traditional ordinary least squares (OLS) regression models. In the case of rain gauge data contaminated by measurement errors with known statistics (mean, variance, and covariance), interpolation should not reproduce the sample data values at their locations; this situation is not treated in this work. Measurement errors, however, could be filtered out in the interpolation procedure via factorial kriging (Wackernagel 1995; Goovaerts 1997).

We now present two geostatistical alternatives for incorporating several relevant predictors into the mapping of rainfall. In what follows, we do not adopt the geostatistical approach of cokriging (Hevesi et al. 1992), because of space limitations and for the following two reasons: (a) the inference effort required by cokriging increases with the number of predictors and (b) various investigators have reported that cokriging does not im-

prove significantly the accuracy of precipitation estimates when compared (in terms of cross-validation statistics) with the two approaches presented in the next section (Goovaerts 2000).

b. Spatial interpolation accounting for rainfall predictors

A straightforward approach for incorporating the most relevant precipitation predictors into the spatial interpolation of rainfall is multiple regression. Let $\mathbf{Y} = [y_k(\mathbf{u}_\alpha), \alpha = 1, \ldots, n, k = 0, \ldots, K]'$ denote an $[n \times (K + 1)]$ matrix containing the values of the predictors at the n grid cells nearest to the n rain gauges. The kth column of $\mathbf{Y}$ is the vector $\mathbf{y}_k = [y_k(\mathbf{u}_\alpha), \alpha = 1, \ldots, n]'$ with entries being the n values of the kth predictor y_k, with, by convention, $y_0(\mathbf{u}_\alpha) = 1$. The $(K + 1)$ vector $\mathbf{b} = (b_0, \ldots, b_K)'$ of linear regression coefficients associated with the available predictors is derived as (Draper and Smith 1998):

$$\mathbf{b} = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{z}, \quad (9)$$

where $\mathbf{z} = [z(\mathbf{u}_\alpha), \alpha = 1, \ldots, n]'$ denotes the vector of n sample precipitation values recorded at the n rain gauges.

An estimate $z^*_f(\mathbf{u})$ for the unknown precipitation value $z(\mathbf{u})$ at any location $\mathbf{u}$ is constructed as

$$z^*_f(\mathbf{u}) = f[\mathbf{y}(\mathbf{u})] = \mathbf{y}(\mathbf{u})\mathbf{b}, \qquad \mathbf{u} \in D, \quad (10)$$

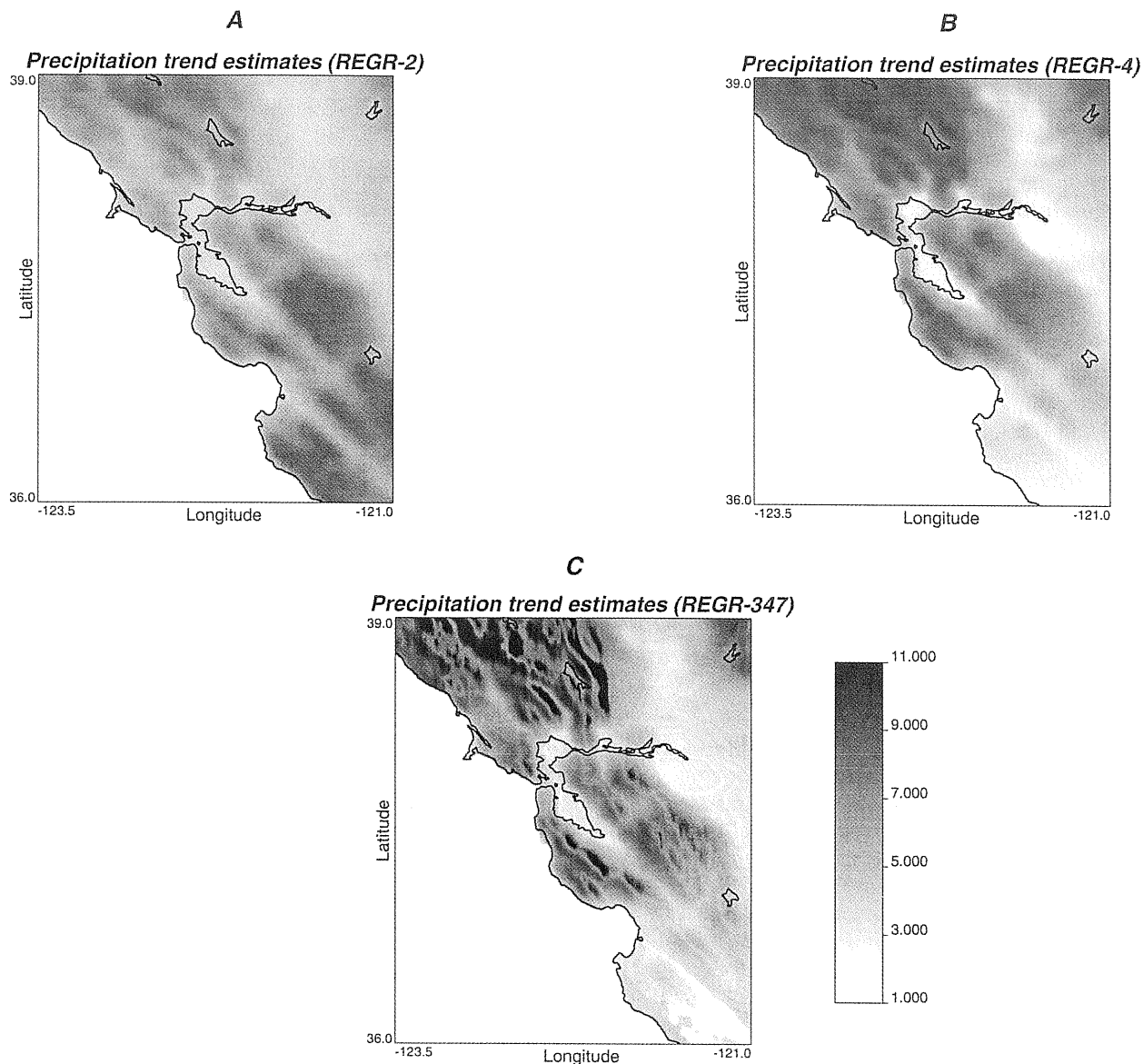where $f( )$ denotes a regression function, and $\mathbf{y}(\mathbf{u}) =$

FIG. 6. Maps of regression-based precipitation estimates (mm) derived using (a) elevation Y_2 as a single predictor (REGR-2); (b) the interaction of specific humidity and elevation Y_4 as a single predictor (REGR-4); and (c) the vertical wind component Y_3, the interaction of specific humidity with elevation Y_4, and the interaction of specific humidity with elevation and vertical wind Y_7 as predictors (REGR-347).

$[y_0(\mathbf{u}), \ldots, y_K(\mathbf{u})]$ denotes the $(K + 1)$ vector of predictor variables available at each map location $\mathbf{u}$, with, by convention, $y_0(\mathbf{u}) = 1$.

The coefficient vector $\mathbf{b}$ is constant over the study region, because the regression function is determined using all n data values; this implies that the function $f(\ )$ does not account for local variations in the relationship between precipitation and its predictors. Any difference (spatial variation) between two estimates $z_f^*(\mathbf{u})$ and $z_f^*(\mathbf{u}')$ at two locations $\mathbf{u}$ and $\mathbf{u}'$ stems solely from the difference of the corresponding predictor vectors $\mathbf{y}(\mathbf{u})$ and $\mathbf{y}(\mathbf{u}')$ at these locations. The estimate $z_f^*(\mathbf{u})$ accounts for the relevance of the collocated predictor vector $\mathbf{y}(\mathbf{u})$ to the un-

known precipitation value $z(\mathbf{u})$; it does not explicitly account, however, for any nearby sample precipation data $z(\mathbf{u}_\alpha)$ or for any predictor vector $\mathbf{y}(\mathbf{u}')$ at a nearby location $\mathbf{u}'$.

A regression-derived precipitation estimate $z_f^*(\mathbf{u}_\alpha)$ at a rain gauge location $\mathbf{u}_\alpha$ typically differs from the sample precipitation value $z(\mathbf{u}_\alpha)$ at that location by a residual amount that varies from one rain gauge to another. If the method of OLS is used to establish the coefficient vector $\mathbf{b}$, the n residual values $\{r(\mathbf{u}_\alpha) = [z(\mathbf{u}_\alpha) - z_f^*(\mathbf{u}_\alpha)], \alpha = 1, \ldots, n\}$ are assumed to have a Gaussian distribution and a purely random (white noise) spatial variation. This latter assumption of spa-

tial independence reduces the applicability of hypothesis-testing procedures in a spatial setting, given that the residuals are usually correlated in space (see section 4).

Recall that, through the subjective decomposition of the RF model in Eq. (4), the observed value $z(\mathbf{u}_\alpha)$ at any rain gauge location $\mathbf{u}_\alpha$ can be viewed as the sum of a local trend component $m_{\text{alg}}(\mathbf{u}_\alpha)$ and a local residual component $r_{\text{alg}}(\mathbf{u}_\alpha)$. One could choose to identify the local trend component $m_{\text{alg}}(\mathbf{u}_\alpha)$ to the regression-based precipitation estimate $z_f^*(\mathbf{u}_\alpha)$ at that location; that is,

$$z(\mathbf{u}_\alpha) = m_{\text{alg}}(\mathbf{u}_\alpha) + r_{\text{alg}}(\mathbf{u}_\alpha) = z_f^*(\mathbf{u}_\alpha) + r_f(\mathbf{u}_\alpha),$$

$$\mathbf{u} \in D, \tag{11}$$

where $r_f(\mathbf{u}_\alpha)$ denotes the residual from the particular trend model $z_f^*(\mathbf{u}_\alpha)$.

Note that the decision to adopt a linear versus a non-linear regression model for establishing the local trend component $m_{\text{alg}}(\mathbf{u}_\alpha)$ is one possible (subjective) interpretation of the decomposition in Eq. (4). The use of forward stepwise regression instead of backward elimination for selecting the pool of most relevant predictors (Draper and Smith 1998) is similarly another possible (subjective) interpretation of the general decomposition in Eq. (4). Different decisions evidently result into different sets of predictor variables, which in turn result into a different trend component $m_{\text{alg}}(\mathbf{u})$ and consequently a different residual component $r_{\text{alg}}(\mathbf{u})$ at each location $\mathbf{u}$. In any case, the residual component $r_{\text{alg}}(\mathbf{u})$ is considered to be stationary with zero mean, and SK consequently can be used for mapping the residual spatial variation.

In the case of a spatially varying trend component $z_f^*(\mathbf{u})$, SK provides an estimate $r_{\text{SK}}^{*f}(\mathbf{u})$ for the unknown residual $r_f(\mathbf{u})$ based on the $n(\mathbf{u})$ residual values within a neighborhood $W(\mathbf{u})$ centered at $\mathbf{u}$. The information utilized by SK comprises the $n(\mathbf{u})$ residual values $\{r_f(\mathbf{u}_\alpha) = [z(\mathbf{u}_\alpha) - z_f^*(\mathbf{u}_\alpha)], \alpha = 1, \ldots, n(\mathbf{u})\}$ within $W(\mathbf{u})$, as well as their spatial covariance $C_{R_f}(\mathbf{h})$. This estimated residual $r_{\text{SK}}^{*f}(\mathbf{u})$ is then added back to the re-

gression-derived estimates $z_f^*(\mathbf{u})$, and the combined procedure is termed detrended kriging (Switzer 1979; Chua and Bras 1982) or SK with varying local mean (SKLM; Goovaerts 1997). The corresponding SKLM estimate $z_{\text{SKLM}}^*(\mathbf{u})$ is written as

$$z_{\text{SKLM}}^*(\mathbf{u}) = m_f^*(\mathbf{u}) + r_{\text{SK}}^{*f}(\mathbf{u})$$

$$= z_f^*(\mathbf{u}) + \sum_{\alpha=1}^{n(\mathbf{u})} \xi_\alpha(\mathbf{u}) r_f(\mathbf{u}_\alpha), \tag{12}$$

where the weights $\xi_\alpha(\mathbf{u})$ are determined per solution of a SK system similar to that of Eq. (6):

$$\sum_{\beta=1}^{n(\mathbf{u})} \xi_\beta(\mathbf{u}) C_{R_f}(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_{R_f}(\mathbf{u}_\alpha - \mathbf{u}),$$

$$\alpha = 1, \ldots, n(\mathbf{u}), \tag{13}$$

the only difference being that instead of the covariance $C_R(\mathbf{h})$ used in Eq. (6), the covariance $C_{R_f}(\mathbf{h})$ of the new residual component $R_f(\mathbf{u})$ is used here. This is also the reason for the different notation $\xi_\alpha(\mathbf{u})$ adopted to denote the corresponding weights, instead of $\lambda_\alpha(\mathbf{u})$ or $\nu_\alpha(\mathbf{u})$.

Local variations in the relation between precipitation and its predictors can be accounted for a posteriori, that is, by locally deforming the regression-based trend component $m_f^*(\mathbf{u}) = z_f^*(\mathbf{u})$. A new trend component $m^*(\mathbf{u})$ can be defined as a local linear rescaling of the regression-based $z_f^*(\mathbf{u})$ estimate, so that the latter conforms locally [within each neighborhood $W(\mathbf{u})$] to the $n(\mathbf{u})$ nearby sample precipitation data $\{z(\mathbf{u}_\alpha), \alpha = 1, \ldots, n(\mathbf{u})\}$. The locally modified trend component $m_{\text{KED}}^*(\mathbf{u})$ is provided via kriging with an external drift (KED) [for example, Wackernagel (1995)]:

$$m_{\text{KED}}^*(\mathbf{u}) = d_0(\mathbf{u}) + d_1(\mathbf{u}) z_f^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \theta_\alpha(\mathbf{u}) z(\mathbf{u}_\alpha), \tag{14}$$

where $d_0(\mathbf{u})$ and $d_1(\mathbf{u})$ denote local regression coefficients, which are constant within each neighborhood $W(\mathbf{u})$ and are different from one neighborhood to another.

The $n(\mathbf{u})$ weights $\{\theta_\alpha(\mathbf{u}), \alpha = 1, \ldots, n(\mathbf{u})\}$ are obtained per solution of the KED system:

$$\left\{ \begin{array}{l} \displaystyle\sum_{\beta=1}^{n(\mathbf{u})} \theta_\beta(\mathbf{u}) C_{R_f}(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_1^{\text{KED}}(\mathbf{u}) + \mu_2^{\text{KED}}(\mathbf{u}) = 0, \qquad \alpha = 1, \ldots, n(\mathbf{u}) \\[2ex] \displaystyle\sum_{\beta=1}^{n(\mathbf{u})} \theta_\beta(\mathbf{u}) = 1 \\[2ex] \displaystyle\sum_{\beta=1}^{n(\mathbf{u})} \theta_\beta(\mathbf{u}) m_f^*(\mathbf{u}_\beta) = m_f^*(\mathbf{u}), \end{array} \right. \tag{15}$$

where $C_{R_f}(\mathbf{u}_\alpha - \mathbf{u}_\beta)$ denotes the covariance between any two residual values $r_{f'}(\mathbf{u}_\alpha)$ and $r_{f'}(\mathbf{u}_\beta)$, with $r_{f'}(\mathbf{u}_\alpha)$

$= z(\mathbf{u}_\alpha) - m_{\text{KED}}^*(\mathbf{u}_\alpha)$; $\mu_1^{\text{KED}}(\mathbf{u})$ and $\mu_2^{\text{KED}}(\mathbf{u})$ are two Lagrange parameters due to the two constraints on the

weights. As a first approximation, the residual covariance $C_{R_f}(\mathbf{h})$ could be identified to the covariance $C_{R_f}(\mathbf{h})$ of the residual values obtained from the (global) regression. The corresponding KED estimate $z^*_{KED}(\mathbf{u})$ for the unknown value $z(\mathbf{u})$ is:

$$z^*_{KED}(\mathbf{u}) = m^*_{KED}(\mathbf{u}) + r^{*f'}_{SK}(\mathbf{u})$$

$$= m^*_{KED}(\mathbf{u}) + \sum_{\alpha=1}^{n(\mathbf{u})} \xi_\alpha(\mathbf{u}) r_{f'}(\mathbf{u}_\alpha), \quad (16)$$

which, as in the previous section, amounts to adding to the KED-derived trend component $m^*_{KED}(\mathbf{u})$ an SK estimate $r^{*f'}_{SK}(\mathbf{u})$ of the corresponding unknown residual $r_{f'}(\mathbf{u})$. Because $C_{R_f}(\mathbf{h}) \simeq C_{R_f}(\mathbf{h})$, the same notation $\xi_\alpha(\mathbf{u})$ is used for the SK weights in SKLM [Eq. (12)] and KED [Eq. (16)].

An alternative (and more elaborate) procedure could be envisaged, whereby the local trend of precipitation is evaluated with respect to each predictor individually instead of collectively through their linear combination obtained by regression. For the case of KED, this approach would call for the determination of two coefficients for each variable, leading to a total of $2(K + 1)$ trend coefficients for the case of $(K + 1)$ variables. In this paper, we do not pursue this alternative; we essentially regard the regression-based precipitation predictions as a secondary variable carrying all the available information content of the individual predictors.

The different geostatistical algorithms presented in this section for mapping precipitation are a consequence of the different decompositions of the RF model $\{Z(\mathbf{u}), \mathbf{u} \in D\}$ given in Eq. (3). To be more specific, their difference lies in the complexity of the spatial trend component $m(\mathbf{u})$ within each neighborhood $W(\mathbf{u})$ centered at the location $\mathbf{u}$ where estimation is performed and the complexity of the resulting residual component. For SK, the spatial trend component $m(\mathbf{u})$ is assumed to be constant; that is, $m(\mathbf{u}) = m, \forall \mathbf{u} \in D$. This assumption is oversimplified, and SK is not considered in what follows. For OK, $m(\mathbf{u})$ is estimated using only nearby precipitation data based on a model of their spatial correlation; that is, $m(\mathbf{u}) = m^*_{OK}(\mathbf{u})$. For SKLM, $m(\mathbf{u})$ is identified to the precipitation estimates based on the regression between lower-atmosphere state variables (including their interactions with local terrain) and observed precipitation; that is, $m(\mathbf{u}) = z^*_f(\mathbf{u})$. Last, for KED, $m(\mathbf{u})$ is identified to local deformations of $z^*_f(\mathbf{u})$; that is, $m(\mathbf{u}) = m^*_{KED}(\mathbf{u})$.

Recall that the resulting residual component $R(\mathbf{u})$ is considered to have a constant (zero) mean and a stationary covariance model $C_R(\mathbf{h})$ in all forms of kriging presented in this section. As a consequence, it is expected that the introduction of more realism in the spatial trend component, via consideration of relevant predictor variables, will render the assumption of a stationary residual less unrealistic. This implies that SKLM and KED, whose spatial trend components are the most

complex, will lead to more realistic (physically less inconsistent) rainfall maps.

Both SKLM and KED can be viewed as corrections to the "first-guess" field derived via an OLS regression. Recall that OLS assumes that the regression residuals are spatially uncorrelated; hence SKLM and KED can be regarded as procedures for reintroducing such a residual spatial correlation to the mapping of precipitation. Both SKLM and KED corrections ensure data exactitude and could be viewed as (intermittent) data assimilation techniques (Rutherford 1972). The only difference is that in the meteorological definition of data assimilation, the first-guess field is a physically based prediction from a dynamic atmospheric model. In our case, the first-guess field is a statistical prediction based on relevant (in a physical sense) variables with observed precipitation. By construction, the regression-based estimates are independent of the resulting residuals, whereas this might not be true in the case of data assimilation involving the difference between atmospheric model predictions and observed precipitation.

The alternative geostatistical algorithms (apart from SK) presented in this section are subsequently applied for mapping rainfall over the study area shown in Fig. 1a.

4. Case study

The precipitation dataset shown in Fig. 1a, as well as the predictors derived in section 2 and shown in Figs. 3 and 4, are used for mapping rainfall over a 300×360 grid of cell size 1 km². In what follows, we first present maps of rainfall estimates derived from the different algorithms presented in section 3. These estimates then are compared in terms of their cross-validation statistics in section 4b. All geostatistical analyses were performed using the public-domain "GSLIB" geostatistical software library package (Deutsch and Journel 1998).

a. Regression models between precipitation and its predictors

In this work, we investigate three possible subsets of predictors from the full set shown in Table 2, that is, three different matrices $\mathbf{Y}_i$ in Eq. (9), to arrive at three different regression coefficient vectors $\mathbf{b}_i$, each specific to the ith predictor subset. For the first regression model, we follow the traditional avenue and use elevation as the single precipitation predictor; that is $\mathbf{Y}_1 = [\mathbf{1} \ \mathbf{y}_2]$, where $\mathbf{1}$ denotes the $n \times 1$ unit vector and $\mathbf{y}_2$ denotes the $n \times 1$ vector of elevation values at the pixels closest to the $n = 77$ rain gauge locations. The resulting regression equation is

$$z^*_2(\mathbf{u}) = 3.7515 + 4.5864 y_2(\mathbf{u}), \quad (17)$$

where the subscript 2 implies that only elevation $y_2(\mathbf{u})$ is used as precipitation predictor. For this regression model, variance $R^2 = 0.12$, indicating that only 12% of
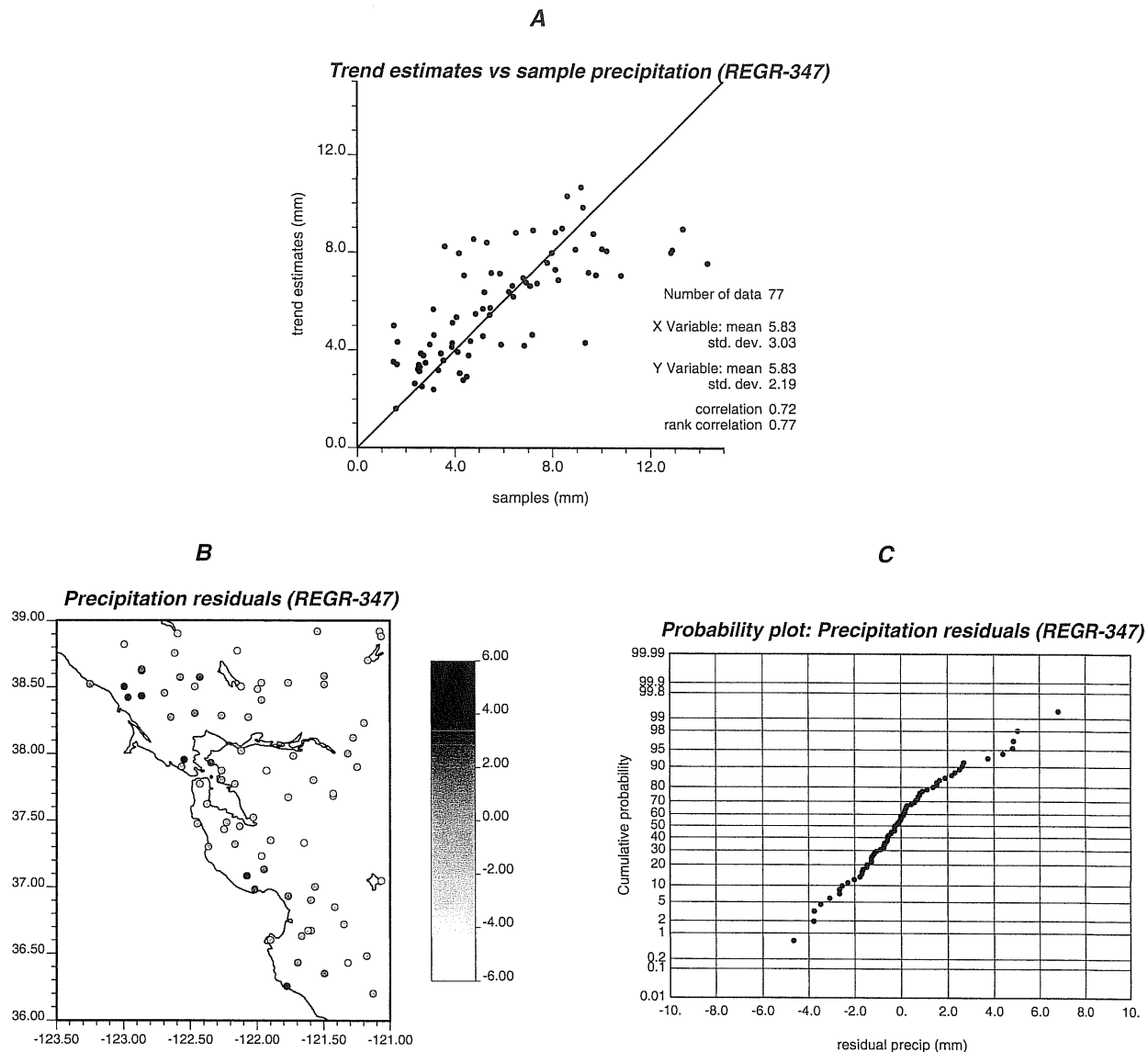
A

Trend estimates vs sample precipitation (REGR-347)

Number of data 77

X Variable: mean 5.83
std. dev. 3.03

Y Variable: mean 5.83
std. dev. 2.19

correlation 0.72
rank correlation 0.77

B   *C*

Precipitation residuals (REGR-347)

Probability plot: Precipitation residuals (REGR-347)

FIG. 7. Characteristics of estimated and residual precipitation values (mm) resulting from regression using the vertical wind component Y_3, the interaction of specific humidity with elevation Y_4, and the interaction of specific humidity with elevation and vertical wind Y_7 as predictors (REGR-347): (a) scatterplot of estimated versus sample precipitation values, (b) location map of precipitation residuals, and (c) normal probability plot of the distribution of precipitation residuals, indicating a quasi-Gaussian distribution.

the precipitation spatial variability is "explained" by elevation; the rmse is 2.88 mm.

The map of precipitation estimates $z_2^*(\mathbf{u})$ derived via Eq. (17) is shown in Fig. 6a. Note the smooth spatial characteristics of this regression-based field, which does not account for either orographic lifting or for humidity. The terrain pattern is imprinted on the map of regression estimates, but there is no rain shadow on the leeward slopes of the terrain nor any indication of more intense precipitation on the windward side (recall that the prevailing wind direction is from the west).

Second, we use the predictor that is best correlated with observed precipitation; that is, we use the inter-

action of specific humidity with elevation as the single precipitation predictor leading to $\mathbf{Y}_2 = [\mathbf{1} \ \mathbf{y}_4]$. The resulting regression equation is

$$z_4^*(\mathbf{u}) = 1.7493 + 7.8493 y_4(\mathbf{u}), \qquad (18)$$

where the subscript 4 implies that only the interaction of specific humidity with elevation $y_4(\mathbf{u})$ is used as precipitation predictor. For this regression model, $R^2 = 0.48$, a 300% increase from the model in Eq. (17), and rmse = 2.22 mm, a 23% decrease from Eq. (17).

The map of precipitation estimates $z_4^*(\mathbf{u})$ derived via Eq. (18) is shown in Fig. 6b. The regression-based field depicts higher precipitation values where elevation is

higher and moisture is more available. The most pronounced difference between the patterns in Fig. 6a and Fig. 6b is found in the southeastern part of the study region, where the model in Eq. (18) yields low precipitation values. Although elevation is high in this area, specific humidity is low, resulting in less precipitation.

Last, we use three predictors that are determined as most significant using backward elimination. Backward elimination is a variable selection procedure whereby one starts with the entire set of predictors and subsequently drops the least significant variable (according to an F test) from this initial set. Variables are repeatedly eliminated from the initial set until no remaining variable can be dropped (again according to the F test). For details regarding backward elimination, and variable selection in general, the reader is referred to Draper and Smith (1998). In this way, we retained the vertical wind component, the interaction of specific humidity with elevation, and the interaction of specific humidity with elevation and vertical wind as the three most important predictors leading to $\mathbf{Y}_3 = [\mathbf{1} \ \mathbf{y}_3 \ \mathbf{y}_4 \ \mathbf{y}_7]$. The resulting regression equation is

$$z^*_{347}(\mathbf{u}) = 0.3862 + 4.8636 y_3(\mathbf{u}) + 13.3508 y_4(\mathbf{u})$$
$$- 7.5561 y_7(\mathbf{u}), \tag{19}$$

where the subscript 347 implies that the predictors $y_3(\mathbf{u})$, $y_4(\mathbf{u})$, and $y_7(\mathbf{u})$ were used in the regression model. For this regression model, $R^2 = 0.52$, a 333% increase from model (17), and rmse = 2.14 mm, a 26% reduction from Eq. (17).

Recall from Table 2 that, although the vertical wind component Y_3 is not correlated with the interaction between humidity and elevation Y_4 ($\rho_{Y_3 Y_4} = 0.03$), the all-interactions term Y_7 is significantly correlated with both Y_3 and Y_4 ($\rho_{Y_7 Y_3} = 0.61$ and $\rho_{Y_7 Y_4} = 0.75$, respectively). The regression coefficients of Eq. (19) therefore should be interpreted in the light of multicollinearity. In particular, the negative sign of the last regression coefficient ($b_4 = -7.55$) is a consequence of multicollinearity. In other words, the influence of the all-interactions predictor Y_7 is negative, once the effect of the other predictors Y_3 and Y_4 is removed. This effect could be also quantified in terms of the corresponding coefficient of partial determination. Note that principal components regression (Draper and Smith 1998), that is, multiple regression using orthogonal (uncorrelated) projections of the original correlated variables, is an alternative procedure for variable selection and regression building when multicollinearity is present. We did not pursue this alternative, because it calls for equally subjective decisions as backward elimination and because the interpretation of the resulting regression coefficients (not in terms of the orthogonal projections but in terms of the original predictors) is as difficult as that of Eq. (19).

The map of precipitation estimates $z^*_{347}(\mathbf{u})$ derived via Eq. (19) is shown in Fig. 6c. The large-scale precipitation patterns of this regression-based field are similar to those of Fig. 6b, but the small-scale patterns differ locally and are more closely associated with terrain variations. In particular, there are obvious rain shadows on the leeward slopes of the terrain (recall that the prevailing wind direction is from the west).

The correlation coefficient between regression-based and observed precipitation values for the regression model in Eq. (19) is 0.72 (see Fig. 7a). Such correlation values between regression-based and rain gauge data are comparable with those obtained from predictions using deterministic regional models with limited inclusion of atmospheric physics and dynamics; for this latter case, such correlation values range between 0.61 and 0.84 (Alpert and Shafir 1989a; Sinclair 1994). The spatial organization of the resulting residual values $\{r_{347}(\mathbf{u}_\alpha) = [z(\mathbf{u}_\alpha) - z^*_{f347}(\mathbf{u}_\alpha)], \alpha = 1, \ldots, n\}$ is shown in Fig. 7b. The normal probability plot of these r_{347} values (Fig. 7c) indicates a quasi-Gaussian distribution. Note that among all precipitation residuals resulting from Eqs. (17)–(19), the r_{347} values exhibit the smallest spatial correlation (see next section), and their histogram is the closest to a Gaussian distribution.

The three different maps of Fig. 6 represent three different first-guess fields, that is, three different interpretations of the relative importance of the seven predictor variables available over the study area. We now proceed by incorporating these three different maps, as well as the information brought by the precipitation residuals at the rain gauge stations (and their spatial correlation), into the geostatistical mapping of rainfall.

b. Geostatistical mapping of rainfall

All geostatistical interpolation procedures presented in the previous section call for a covariance (or, equivalent, a variogram) function that models the spatial continuity (or, equivalent, the spatial variability) of the original sample precipitation z data or that of the residual r values resulting from the particular trend function adopted. Figure 8 depicts the omnidirectional experimental and model variograms of the sample precipitation data (Fig. 8a), as well as those corresponding to the residuals from different trend functions (Figs. 8b–d).

Two nested variogram structures (with no nugget effect), corresponding to one small-scale and one large-scale spatial process, were used to model the sample variograms of the ith dataset, be it the original rain gauge z data or any set of residual r values:

$$\gamma^{(i)}(|\mathbf{h}|) = C^{(i)}(\mathbf{0}) - c_1^{(i)} \exp\left[\frac{3|\mathbf{h}|}{a_1^{(i)}}\right]$$
$$- c_2^{(i)} \exp\left[\frac{3|\mathbf{h}|}{a_2^{(i)}}\right], \tag{20}$$

where $|\mathbf{h}|$ denotes the modulus of a distance vector $\mathbf{h}$, and $C^{(i)}(\mathbf{0})$ denotes the variance of the ith dataset. Pa-
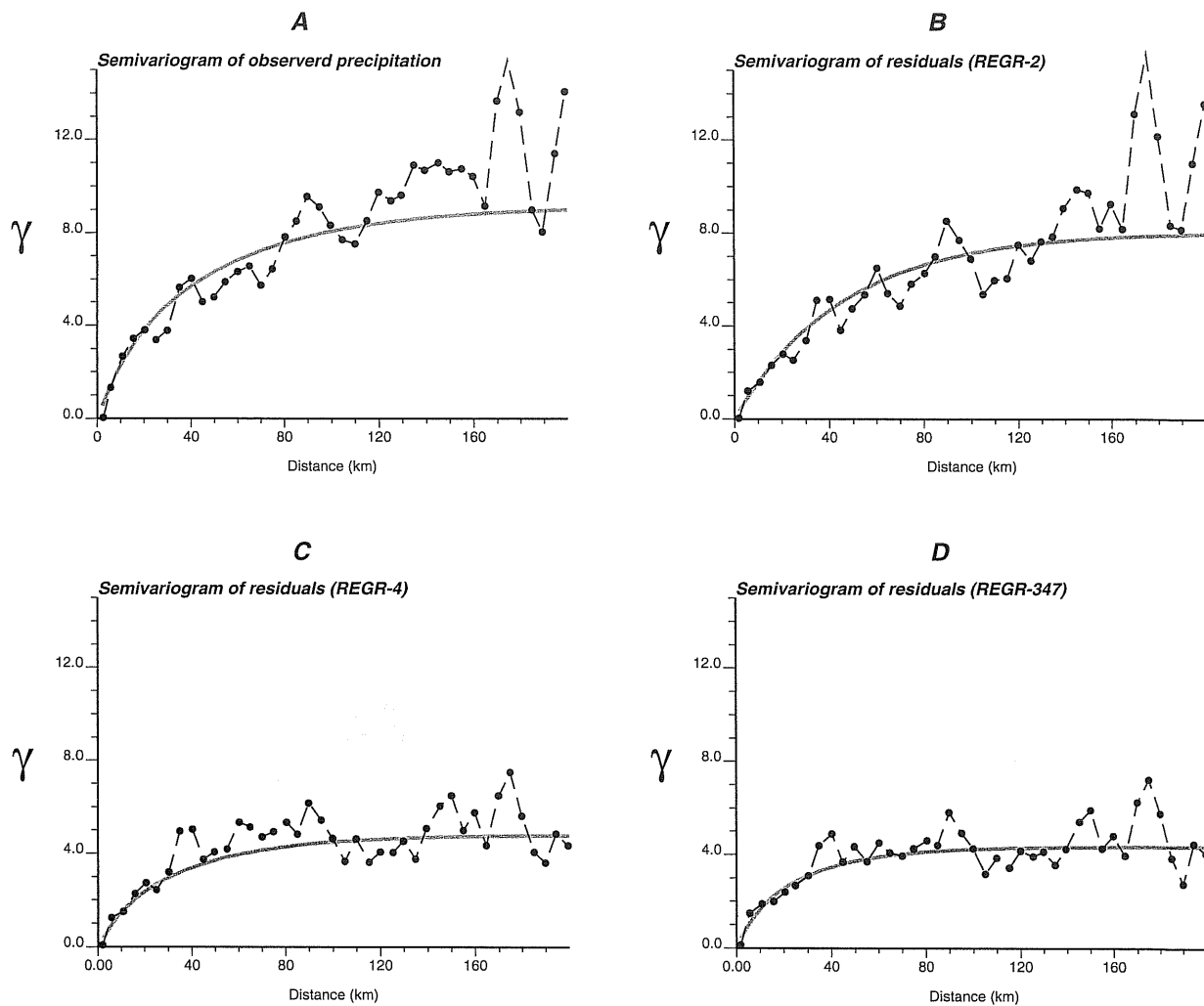
FIG. 8. Spatial variability characteristics of observed and residual datasets at the 77 rain gauges. Sample (thick dash–dotted black lines) and model (thick gray lines) variograms for (a) observed precipitation; (b) residuals from regression estimates using elevation Y_1 as the only predictor (REGR-2); (c) residuals from regression estimates using the interaction of specific humidity and elevation Y_4 as the only predictor (REGR-4); and (d) residuals from regression estimates using vertical wind Y_3, the interaction of specific humidity with elevation Y_4, and the interaction of specific humidity with elevation and vertical wind Y_7 as the vector of predictors (REGR-347).

rameters $c_1^{(i)}$ and $a_1^{(i)}$ denote the respective sill (contribution to the total variance of the ith dataset) and range of the small-scale variogram structure; $c_2^{(i)}$ and $a_2^{(i)}$ denote the respective sill and range of the large-scale variogram structure.

The parameters of the variogram models for the original precipitation z data and the residual r values from various trend functions were derived via cross-validation [see Isaaks and Srivastava (1989) and next section] and are tabulated in Table 3. Because a zero nugget effect was adopted for all cases, the sum of the sills of the two variogram structures should be approximately equal to the variance of the respective dataset. In the case of original rain gauge z samples, precipitation spatial variability is decomposed in a small-scale (40 km) process that explains 22% of the total variance (=9.18 mm^2), and a large-scale (160 km) process that explains

the remaining 78% of the total variance. Similar decompositions, although with different parameters, were adopted for the residual r datasets. In general, as the trend function becomes more complex, thus involving more predictor variables and explaining a larger proportion of the z sample variability, the variance of the corresponding residual r values [sum of $c_1^{(i)}$ and $c_2^{(i)}$] becomes smaller and the corresponding nested variogram structures exhibit smaller ranges [$a_1^{(i)}$ and $a_2^{(i)}$]; see Table 3.

The map of precipitation estimates $z_{OK}^*(\mathbf{u})$ derived via OK using Eqs. 7 and 8 is shown in Fig. 9. Note the smooth pattern of spatial variability exhibited by the OK precipitation estimates and the reproduction of large-scale features found in the rain gauge map of Fig. 1a. A rain shadow on the leeward slopes of the terrain is hardly discernible in this case. The map of precipitation estimates $z_{SKLM_2}^*(\mathbf{u})$

A

OK precipitation estimates

B

SKLM-2 precipitation estimates

C

SKLM-4 precipitation estimates
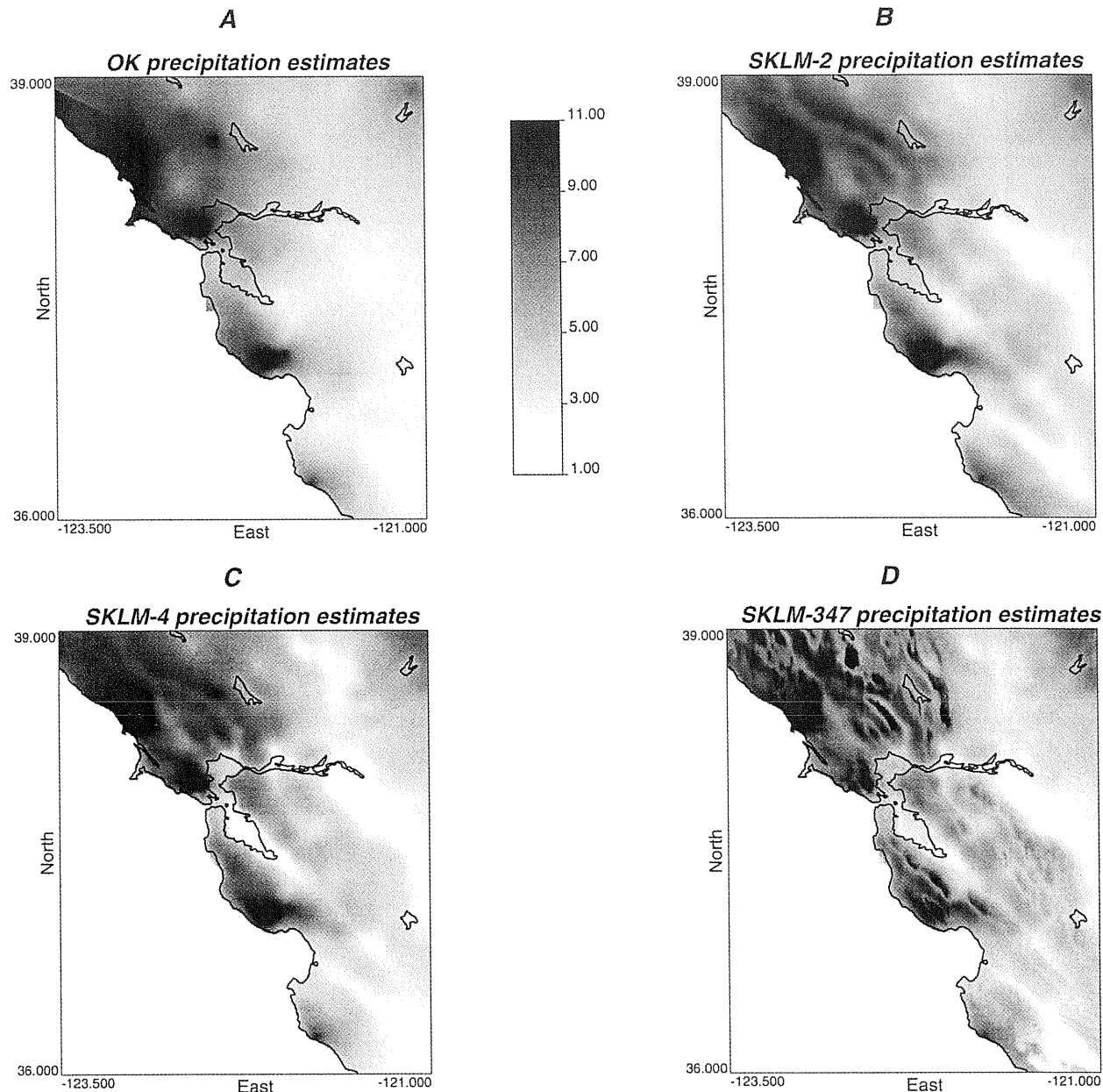

D

SKLM-347 precipitation estimates

FIG. 9. Kriging-derived precipitation estimates (mm) based on (a) only station data and ordinary kriging (OK); (b) regression using elevation Y_2 as the single precipitation predictor, followed by simple kriging of residuals (SKLM-2); (c) regression using the interaction between specific humidity and elevation Y_4 as the single precipitation predictor, followed by simple kriging of residuals (SKLM-4); and (d) regression using the vertical wind component Y_3, the interaction between specific humidity and elevation Y_4, and the interaction among specific humidity, elevation, and the vertical wind component Y_7 as three precipitation predictors, followed by simple kriging of residuals (SKLM-347).

derived via simple kriging with locally varying mean using Eqs. 12 and 13, with the regression-based trend component $z_2^*(\mathbf{u})$ as local mean, is shown in Fig. 9b. The pattern of spatial variability in the resulting estimates of Fig. 9b is now less smooth than that of the OK estimates of Fig. 9a. Note that the underlying trend component of Fig. 6b can be distinguished in the map of Fig. 9b. The map of precipitation estimates $z_{\mathrm{SKLM}_4}^*(\mathbf{u})$ derived via SKLM, with the

regression-based trend component $z_4^*(\mathbf{u})$ as local mean, is shown in Fig. 9c. Note the more complex spatial pattern imposed by the corresponding trend component of Fig. 6c. Last, the map of precipitation estimates $z_{\mathrm{SKLM}_{347}}^*(\mathbf{u})$ derived via SKLM, with the regression-based trend component $z_{347}^*(\mathbf{u})$ as local mean, is shown in Fig. 9d. The spatial distribution of the resulting estimates of Fig. 9c exhibits the most complex spatial patterns, especially at

A

KED-2 precipitation estimates

B

KED-4 precipitation estimates
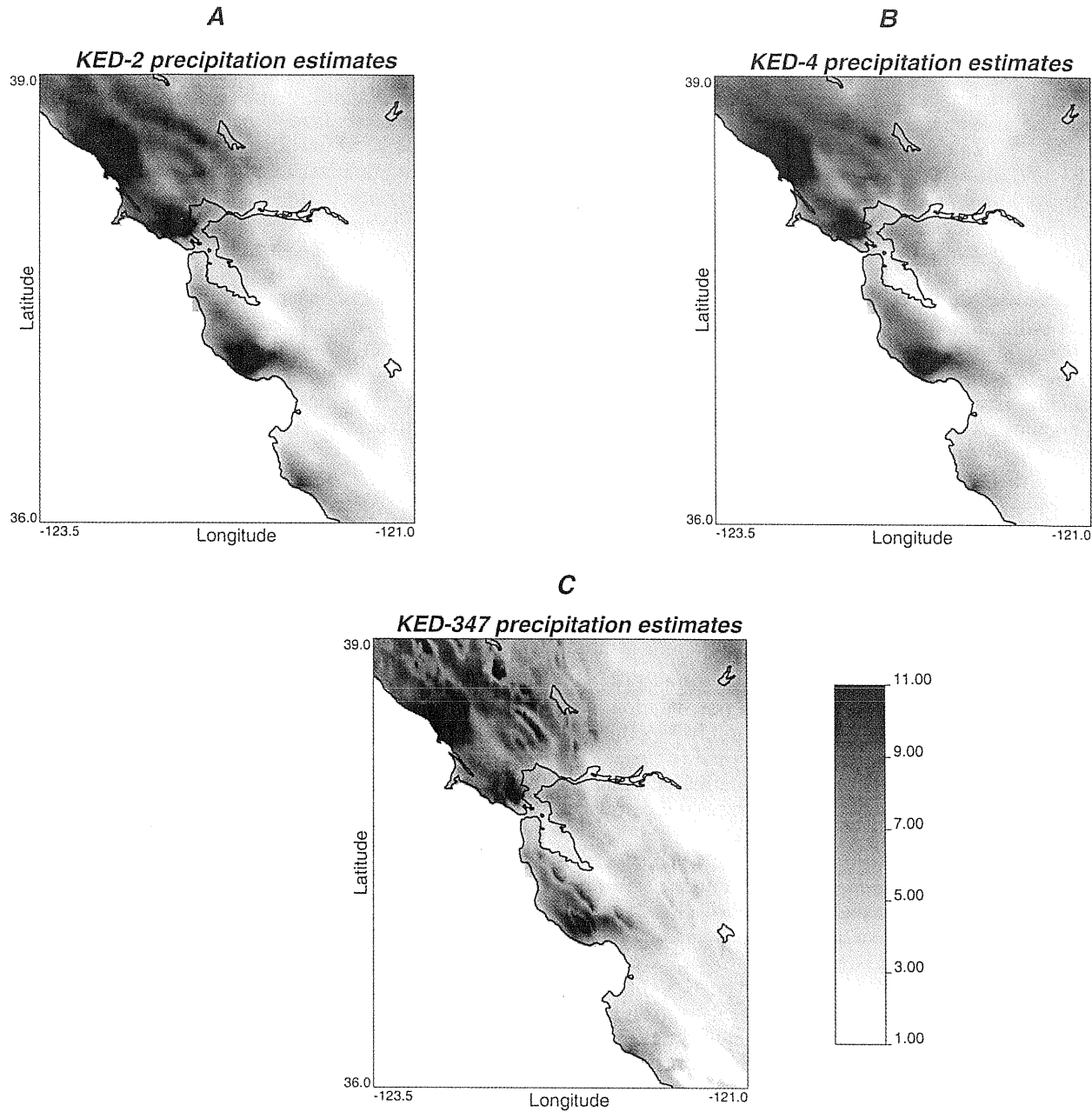

C

KED-347 precipitation estimates

FIG. 10. Kriging-derived precipitation estimates (mm) based on (a) local deformation of the regression-based trend component $z_2^*(\mathbf{u})$, with elevation being used as the single precipitation predictor, followed by simple kriging of residuals (KED-2); (b) local deformation of the regression-based trend component $z_4^*(\mathbf{u})$, with the interaction between specific humidity and elevation being used as the single precipitation predictor, followed simple kriging of residuals (KED-4); and (c) local deformation of the regression-based trend component $z_{347}^*(\mathbf{u})$, with the vertical wind Y_3, the interaction between specific humidity and elevation Y_4, and the interaction among specific humidity, elevation, and vertical wind Y_7 being used as precipitation predictors, followed simple kriging of residuals (KED-347).

small scales. Note the strong rain shadow on the leeward slopes of the terrain, in accordance with the corresponding trend component of Fig. 6d.

The maps of precipitation estimates $z_{\mathrm{KED}}^*(\mathbf{u})$ derived by local deformation of the regression-based trend component $z_f^*(\mathbf{u})$ via KED using Eqs. 16 and 15 are shown in Fig. 10. The trend components $z_f^*(\mathbf{u})$ were those depicted in

Fig. 6D and used for deriving the SKLM estimates of Fig. 9. The residual covariance model adopted was identified to that of the r' residuals; that is, $C_{R_{f'}}(\mathbf{h}) \simeq C_{R_f}(\mathbf{h})$. The pattern of spatial variability is as complex as that found in the SKLM estimates of Fig. 9, in accordance with the corresponding trend components of Fig. 6, yet local differences between the corresponding SKLM and KED

maps can be distinguished (compare, for example, the north-central parts of Figs. 9d and 10c); it is likely that KED leads to somewhat less variable precipitation estimates in areas away from the rain gauges.

The task now is to compare objectively these alternative rainfall maps and to evaluate the improvement (if any) brought by the predictors to the accuracy of the derived map product.

c. Comparison of alternative mapping procedures

Because all variants of kriging are exact interpolators, no estimation error occurs at rain gauge locations. The different geostatistical interpolation procedures are therefore compared via cross-validation [see, for example, Isaaks and Srivastava (1989)]. Cross-validation amounts to sequentially dropping a single precipitation value from the sample dataset and reestimating its value from the remaining samples using all other available information. At any sample location $\mathbf{u}_\alpha$, both the original precipitation value $z(\mathbf{u}_\alpha)$ and its cross-validation-derived estimate $z_{cr}^*(\mathbf{u}_\alpha)$ are available. Comparison of the various approaches therefore can be performed by examining the statistics and spatial patterns of the 77 cross-validation error values $\{e_{cr}^*(\mathbf{u}_\alpha) = [z_{cr}^*(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha)], \alpha = 1, \ldots, n\}$ computed at the 77 rain gauge locations. A positive cross-validation error $e_{cr}^*(\mathbf{u}_\alpha)$ indicates overestimation of the actual precipitation $z(\mathbf{u}_\alpha)$ by the cross-validation estimate $z_{cr}^*(\mathbf{u}_\alpha)$, whereas a negative cross-validation error indicates the reverse.

Table 4 gives selected cross-validation statistics for the interpolation algorithms considered in the previous section. Cross-validation statistics examined are the rmse, as well as the correlation $\rho_{Z_{cr}^*Z}$ between the cross-validation estimates and the true (sample) values and the correlation $\rho_{E_{cr}^*Z}$ between the cross-validation errors and the true (sample) values. For the rmse, the best case corresponds to near-zero values. For $\rho_{Z_{cr}^*Z}$, the best case is a value of 1, indicating an almost perfect correlation between cross-validation estimates and true values. For $\rho_{E_{cr}^*Z}$, the best case is a value of 0 indicating a lack of correlation between cross-validation errors and true data; this implies that there is no systematic underestimation or overestimation of high or low precipitation values. In Table 4, values in parentheses denote the relative difference in the corresponding cross-validation scores from the case of OK using only rain gauge data.

Ordinary Kriging is the least accurate, because it leads to the largest rmse and $\rho_{E_{cr}^*Z}$ values and to the smallest $\rho_{Z_{cr}^*Z}$ value, of all the algorithms considered (see Table 4). The best performance is achieved by SKLM$_{347}$ and KED$_{347}$, because they lead to the lowest rmse and $\rho_{E_{cr}^*Z}$ values and to relatively high $\rho_{Z_{cr}^*Z}$ values. Note also that SKLM$_{347}$ yields cross-validation errors with the smallest spatial correlation. In other words, out of all algorithms considered, SKLM$_{347}$ yields cross-validation residuals with the smallest variogram range (not shown). For OK, the high negative correlation value $\rho_{E_{cr}^*Z} =$ -0.54 implies that negative (or positive) OK cross-validation errors are systematically associated with high (or low) precipitation values, which is a direct consequence of the smooth trend component $m_{OK}^*(\mathbf{u})$. Low precipitation values tend to be systematically overestimated (leading to positive cross-validation errors), whereas high precipitation values tend to be systematically underestimated (leading to negative cross-validation errors). Such a smoothing effect is much less severe for SKLM$_{347}$ ($\rho_{E_{cr}^*Z} = -0.37$, a 31% reduction from OK) and for SKLM$_4$ ($\rho_{E_{cr}^*Z} = -0.44$, a 15% reduction from OK).

From Table 4, one can conclude that, overall, rainfall mapping via SKLM and KED, accounting for the maximum amount of relevant predictors and for spatial correlation of the resulting residuals, yields relatively better cross-validation scores for the particular dataset and study region. Although OK performs worst in terms of cross-validation scores, its performance is not dramatically different than the other algorithms considered in this work. Such relatively good cross-validation scores for the case of OK are largely due to the rain-gauge density and the large spatial correlation range (160 km) of precipitation.

To arrive at a more definite conclusion regarding the improvement brought by the lower-atmosphere state variables and terrain information, we compare their relationship with the various cross-validation error datasets that result from the different algorithms adopted. In particular, we investigate whether the entire set of available predictors (Table 2) could account for a portion of the spatial variability of the cross-validation errors. If, for example, a statistically significant regression function can be established between the predictors and the cross-validation errors, this would imply that such errors could be potentially reduced, had those predictors been included in the interpolation procedure. Relations between predictors and cross-validation errors were investigated for all the algorithms presented in the previous section.

The regression characteristics between cross-validation errors corresponding to different mapping algorithms and three precipitation predictors Y_3, Y_4, and Y_7 are shown in Table 5. A relatively high R^2 indicates that a higher proportion of the spatial variability of cross-validation errors can be accounted for by the three predictors. A relatively high F statistic associated with a small significance p value implies that the regression model between the corresponding cross-validation errors and the precipitation predictors is statistically significant.

From Table 5, one can see that statistically significant regression models between cross-validation errors and the three precipitation predictors Y_3, Y_4, and Y_7 can be established for the case of OK, SKLM$_2$, and, to a lesser extent, for the case of SKLM$_4$. For the case of OK, the percentage of variance of cross-validation errors accounted for by the regression on the predictors is 24%,

TABLE 5. Regression characteristics between cross-validation errors of different mapping algorithms (subscripts denote the predictors used in the respective regression equations), and the three precipitation predictors: vertical wind component Y_3, interaction of specific humidity with elevation Y_4, and interaction among humidity, elevation, and vertical wind Y_7. Here, R^2 denotes the proportion of cross-validation errors that is explained by regression using the three predictors Y_3, Y_4, and Y_7. A relatively high F statistic associated with a p value smaller than 0.001 implies that the regression model between the corresponding cross-validation errors and the precipitation predictors is statistically significant.

Statistics	OK	$SKLM_2$	KED_2	$SKLM_4$	KED_4	$SKLM_{347}$	KED_{347}
R^2	0.24	0.20	0.19	0.19	0.18	0.04	0.01
F score	7.60	6.17	5.74	5.67	5.49	1.05	0.30
p value	0.0002	0.0009	0.0014	0.0015	0.0019	0.3767	0.8250

whereas such a proportion drops to 19% in the case of $SKLM_4$. Note that KED and SKLM give similar results in all cases. One could argue that the R^2 values of Table 5 for the case of OK, $SKLM_2$, and $SKLM_4$ (or, equivalent, for the corresponding KED algorithms) are not very high. They are, however, important in that the three predictors Y_3, Y_4, and Y_7 can explain a nonnegligible proportion of the spatial variability of the cross-validation errors. Such results corroborate the importance of including lower-atmosphere state variables and their interaction with local terrain characteristics into the spatial interpolation of rainfall.

Last, we evaluate the performance of the various algorithms in terms of jackknife scores. To be more specific, we exclude sample precipitation values at the 15 stations marked with crosses in Fig. 1a and perform estimation at these 15 locations. We preferentially exclude the highest precipitation values for the jackknife, because accurate estimation of such values is critical in many hydrologic analyses. Most of the jackknife locations in the northwestern part of the study region are located outside the convex hull of the remaining rain gauges. This means that estimation at these jackknife locations is performed in extrapolation mode, in which case the local trend model $m_{alg}(\mathbf{u})$ adopted is of paramount importance. The algorithm that will suffer most from this extrapolation setting is OK, because its local trend model is estimated from nearby data that are located (south)east from the jackknife stations (see Fig. 1a). This adverse setting for OK, however, is compensated by the fact that these highest sample precipitation values are not predicted as well as the other ones from the corresponding regression model (local trend; see Fig. 7a). It also should be noted that, for all the algorithms considered, we build the corresponding regression models and infer the resulting residual variograms using all $n = 77$ rain gauges. The resulting rmse of jackknife errors is shown in Table 6. Ordinary kriging yields the largest rmse value (4.57), whereas $SKLM_{374}$ and $SKLM_4$ yield the smallest rmse values (3.42 and

3.69, respectively), thus achieving a 25% and 19% reduction from the OK-based rmse. Again, the importance of including lower-atmosphere state variables and their interaction with local terrain characteristics into the spatial interpolation of rainfall is corroborated.

5. Discussion and conclusions

A geostatistical framework for enhanced analyses of precipitation is presented in this paper. Atmospheric and terrain characteristics, which control the spatial distribution of precipitation at regional scales, are accounted for via alternative forms of kriging. Lower-atmosphere state variables include specific humidity and horizontal wind components, readily available at coarse resolution (2.5° × 2.5°) from the NCEP–NCAR reanalysis products. Their interactions with terrain, both elevation and its local gradients, provide valuable information for mapping the spatial distribution of orographic precipitation. The relevance of this information is first evaluated via a regression model based on collocated precipitation and predictor data. The regression-based precipitation estimates constitute a first-guess field. Spatial interpolation of the residuals from this first-guess field is then performed, and the resulting residual field is added to the regression-based estimates. As an alternative, the first-guess field is locally modified to conform to nearby sample precipitation data, followed by spatial interpolation of the resulting residuals.

The alternative geostatistical procedures, which differ in the complexity of the first-guess field, are used for mapping time-averaged precipitation from a set of 77 rain gauges over a region of northern California for NDJ of 1981/82. For this particular study area and time period, elevation alone explains 12% of the precipitation spatial variability, whereas the interaction of specific humidity with elevation explains 48% of such variability. Linear regression using the vertical wind component, the interaction of specific humidity with elevation, and the interaction of specific humidity with

TABLE 6. Root-mean-square error (mm) of jackknife errors for different mapping algorithms (subscripts denote the predictors used in the respective regression equations). The numbers in parentheses show the relative changes in the corresponding jackknife statistics from OK.

Statistics	OK	$SKLM_2$	KED_2	$SKLM_4$	KED_4	$SKLM_{347}$	KED_{347}
Rmse	4.57	4.29	4.30	3.69 (−19%)	4.28	3.42 (−25%)	4.17

elevation and vertical wind (the latter being a measure of orographic uplift of air masses modulated by a surrogate of temperature) as precipitation predictors explains 52% of the variance of observed precipitation. Different first-guess fields are constructed via linear regression using a different number of precipitation predictors. The resulting residuals are correlated in space, with ranges varying from 150 km (for the least complex first-guess field) to 90 km (for the most complex first-guess field); the correlation range for the sample rain gauge precipitation is 160 km. The various interpolation algorithms are compared in terms of (a) their respective cross-validation error statistics, (b) the significance of regression models between such cross-validation errors and precipitation predictors, and (c) their jackknife errors. In all cases, interpolation using only rain gauge data (OK) performed worst, while interpolation using the maximum amount of relevant atmospheric and terrain information ($SKLM_{347}$) resulted in better cross-validation and jackknife scores. The reduction in rmse values from OK obtained via $SKLM_{347}$ ranged from 9% for cross-validation to 25% for jackknife.

Classical objective analysis schemes ignore important relevant information such as humidity and vertical wind and consequently produce oversmooth representations of the spatial distribution of rainfall; such an adverse effect is intensified when the network of rain gauges is sparse. This paper demonstrates the capability of constructing realistic analyses of precipitation by integrating readily available and physically relevant predictors. Precipitation analyses derived from the proposed schemes can be used as reference for comparison against spatial patterns of precipitation obtained via detailed atmospheric models operating at regional scales.

In conclusion, it should be noted that rainfall mapping within the proposed geostatistical framework could be enhanced by the availability of wind fields at less coarse resolution than those available from the NCEP reanalysis products. Such finer-resolution wind fields could be obtained, for example, from regional-scale atmospheric models with detailed parameterization of physical and dynamical processes and could allow resolution of the local divergence of air masses in complex terrain. This better-resolved vertical wind component could be better correlated with observed rain gauge precipitation. Similar remarks can be made for other relevant variables, such as temperature and humidity, whose availability at regional scales could be critical for improving the final map product.

REFERENCES

Alpert, P., 1986: Mesoscale indexing of the distribution of orographic precipitation over high mountains. *J. Appl. Meteor.,* **25,** 532–545.

——, and H. Shafir, 1989a: A physical model to complement rainfall normals over complex terrain. *J. Hydrol.,* **110,** 51–62.

——, and ——, 1989b: Meso-γ-scale distribution of orographic precipitation: Numerical study and comparison with precipitation derived from radar measurements. *J. Appl. Meteor.,* **28,** 1105–1116.

Andrieu, H., M. N. French, V. Thauvin, and W. F. Krajewski, 1996: Adaptation and application of a quantitative rainfall forecasting model in a mountainous region. *J. Hydrol.,* **184,** 243–259.

Barros, A. P., and D. P. Lettenmaier, 1993: Dynamic modeling of the spatial distribution of precipitation in remote mountainous areas. *Mon. Wea. Rev.,* **121,** 1195–1214.

Burns, J. I., 1953: Small-scale topographic effects on precipitation in San Dimas experimental forest. *Trans. Amer. Geophys. Union,* **34,** 761–767.

Chilès, J. P., and P. Delfiner, 1999: *Geostatistics: Modeling Spatial Uncertainty.* John Wiley and Sons, 695 pp.

Chua, S. H., and R. L. Bras, 1982: Optimal estimators of mean area precipitation in regions of orographic influence. *J. Hydrol.,* **57,** 23–48.

Cressie, N. A. C., 1993: *Statistics for Spatial Data.* John Wiley and Sons, 900 pp.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical–topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.,* **33,** 140–158.

Deutsch, C. V., and A. G. Journel, 1998: *GSLIB: Geostatistical Software Library and User's Guide.* 2d ed. Oxford University Press, 368 pp.

Dirks, K. N., J. E. Hay, C. D. Stow, and D. Harris, 1998: High-resolution studies of rainfall on Norfolk Island. Part II: Interpolation of rainfall data. *J. Hydrol.,* **208,** 187–193.

Draper, N. R., and H. Smith, 1998: *Applied Regression Analysis.* John Wiley and Sons, 706 pp.

Entekhabi, D., and Coauthors, 1999: An agenda for land surface hydrology research and a call for the second International Hydrological Decade. *Bull. Amer. Meteor. Soc.,* **80,** 2043–2058.

Giorgi, F., and L. O. Mearns, 1991: Approaches to the simulation of regional climate change: A review. *Rev. Geophys.,* **29,** 191–216.

Goovaerts, P., 1997: *Geostatistics for Natural Resources Evaluation.* Oxford University Press, 483 pp.

——, 2000: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.,* **228,** 113–129.

Herman, A., V. B. Kumar, P. A. Arkin, and J. V. Kousky, 1997: Objectively determined 10-day African rainfall estimates created for famine early warning systems. *Int. J. Remote Sens.,* **18,** 2147–2159.

Hevesi, J. A., A. L. Flint, and J. D. Istok, 1992: Precipitation estimation in mountainous terrain using multivariate geostatistics. Part II: Isohyetal maps. *J. Appl. Meteor.,* **31,** 677–688.

Isaaks, E., and R. M. Srivastava, 1989: *An Introduction to Applied Geostatistics.* Oxford University Press, 561 pp.

Isakson, A., 1996: Rainfall distribution over central and southern Israel induced by large-scale moisture flux. *J. Appl. Meteor.,* **35,** 1063–1075.

Journel, A. G., and C. J. Huijbregts, 1978: *Mining Geostatistics.* Academic Press, 600 pp.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kim, J., and S.-T. Soong, 1996: Simulation of a precipitation event in the western United States. *Regional Impacts of Global Climate Change,* S. J. Ghan et al., Eds., Battelle Press, 73–84.

——, N. L. Miller, A. K. Guetter, and K. P. Georgakakos, 1998: River

flow response to precipitation and snow budget in California during the 1994/95 winter. *J. Climate,* **11,** 2376–2386.

——, ——, J. D. Farrara, and S.-Y. Hong, 2000: A seasonal precipitation and stream flow hindcast and prediction study in the western United States during the 1997/98 winter season using a dynamic downscaling system. *J. Hydrometeor.,* **1,** 311–329.

Leung, L. R., M. S. Wigmosta, S. J. Ghan, D. J. Epstein, and L. W. Vail, 1996: Application of a subgrid orographic precipitation/ surface hydrology scheme to a mountain watershed. *J. Geophys. Res.,* **101,** 12 803–12 817.

Miller, N. L., and J. Kim, 1996: Numerical prediction of precipitation and river flow over the Russian River watershed during the January 1995 California storms. *Bull. Amer. Meteor. Soc.,* **77,** 101–105.

Pandey, G. R., D. R. Cayan, and K. P. Georgakakos, 1999: Precipitation structure in the Sierra Nevada of California during winter. *J. Geophys. Res.,* **104,** 12 019–12 030.

Rhea, J. O., 1978: Orographic precipitation model for hydrometeorologic use. Ph.D. dissertation, Department of Atmospheric Science, Colorado State University, 199 pp.

Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.,* **29,** 809–815.

Sinclair, M. R., 1994: A diagnostic model for estimating orographic precipitation. *J. Appl. Meteor.,* **33,** 1163–1175.

Smith, R. B., 1979: The influence of mountains on the atmosphere. *Advances in Geophysics,* Vol. 21, Academic Press, 87–230.

Spreen, W. C., 1947: A determination of the effect of topography upon precipitation. *Trans. Amer. Geophys. Union,* **28,** 285–290.

Switzer, P., 1979: Statistical considerations in network design. *Water Resour. Res.,* **15,** 1712–1716.

Tabios, G., and J. Salas, 1985: A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resour. Bull.,* **21,** 365–380.

Thiébaux, H. J., 1997: The power of duality in spatial–temporal estimation. *J. Climate,* **10,** 567–573.

Wackernagel, H., 1995: *Multivariate Geostatistics.* Springer-Verlag, 256 pp.

Wolfson, N., 1975: Topographical effects on standard normals of rainfall over Israel. *Weather,* **30,** 138–144.

Wotling, G., C. Bouvier, J. Danloux, and J.-M. Fritsch, 2000: Regionalization of extreme precipitation distribution using the principal components of the topographic environment. *J. Hydrol.,* **233,** 86–101.